PCT/GB 95 / 01949

17 AUGUST 1995

The Patent Office

The Patent Office
Cardiff Road
Newport
Gwent
NP9 1RH

I, the undersigned, being an officer duly authorised in accordance with Section 62(3) of the Patents and Designs Act 1907, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

PRIORITY DOCUMENT

Signed M.J. Russell

Dated 1st September, 1995

For official use

9416880.4

23AUG 94A09368 7015 PAT 1 77 UC 25.00

Your reference
HCM/C1037/M

20 AUG 1994

Notes
Please type, or write in dark ink using CAPITAL letters. A prescribed fee is payable for a request for grant of a patent. For details, please contact the Patent Office (telephone 071–438 4700).

Rule 16 of the Patents Rules 1990 the main rule governing the completion and filing of this form.

❷ Do not give trading styles, for example, 'Trading as XYZ company', nationality or former names, for example, 'formerly known as ABC Ltd' as these are not required.

THE PATENT OFFICE 20 AUG 1994 RECEIVED BY POST

Warning
After an application for a Patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977 and will inform the applicant if such prohibition or restriction is necessary. Applicants resident in the United Kingdom are also reminded that under Section 23, applications may not be filed abroad without written permission unless an application has been filed not less than 6 weeks previously in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction revoked.

The Patent Office

Request for grant of a Patent

Form 1/77

Patents Act 1977

❶ Title of invention

1 Please give the title of the invention: Improvements in or relating to binding proteins for recognition of DNA

❷ Applicant's details

❑ **First or only applicant**

2a If you are applying as a corporate body please give:

Corporate name: Medical Research Council

Country *(and State of incorporation, if appropriate)*: Great Britain

2b If you are applying as an individual or one of a partnership please give in full:

Surname

Forenames

2c **In all cases**, please give the following details:

Address: 20 Park Crescent
London

UK postcode *(if applicable)*: W1N 4AL

Country: Great Britain

ADP number *(if known)*: 5960070010

❼ *The answer must be 'No' if:*
- *any applicant is not an inventor*
- *there is an inventor who is not an applicant, **or***
- *any applicant is a corporate body.*

❼ Inventorship

7 Are you (the applicant or applicants) the sole inventor or the joint inventors?

Please mark correct box

Yes [ ] No [x] ➡ ***A Statement of Inventorship on Patents Form 7/77 will need to be filed (see Rule 15).***

❽ *Please supply duplicates of claim(s), abstract, description and drawing(s).*

❽ Checklist

8a Please fill in the number of sheets for each of the following types of document contained in this application.

Continuation sheets for this Patents Form 1/77			
Claim(s)		Description	45
Abstract		Drawing(s)	13

8b Which of the following documents also accompanies the application?

Please mark correct box(es)

Priority documents *(please state how many)*	
Translation(s) of Priority documents *(please state how many)*	
Patents Form 7/77 – Statement of Inventorship and Right to Grant *(please state how many)*	
Patents Form 9/77 – Preliminary Examination/Search	
Patents Form 10/77 – Request for Substantive Examination	

❾ *You or your appointed agent (see Rule 90 of the Patents Rules 1990) must sign this request.*

❾ Request

I/We request the grant of a patent on the basis of this application.

Please sign here ➡

Signed Keith W Nash & Co Agents

Date 19 th August 1994 *(day month year)*

A completed fee sheet should preferably accompany the fee.

Please return the completed form, attachments and duplicates where requested, together with the prescribed fee to either:

❑ ***The Comptroller***
The Patent Office
Cardiff Road
Newport
Gwent
NP9 1RH

or

❑ ***The Comptroller***
The Patent Office
25 Southampton Buildings
London
WC2A 1AY

Title: Improvements in or relating to binding proteins for recognition of DNA

SECTION 1

ABSTRACT We have used two selection techniques to study sequence-specific DNA recognition by the zinc finger, the small, modular DNA-binding mini-domain. We have chosen zinc fingers since they bind as independent modules, and so can be linked together in a peptide designed to bind a predetermined DNA site. In this paper, we describe how a library of zinc fingers displayed on the surface of bacteriophage enables selection of fingers capable of binding to given DNA triplets. The amino acid sequences of selected fingers which bind the same triplet are compared to examine how sequence specific DNA recognition occurs. Our results can be rationalised in terms of coded interactions between zinc fingers and DNA, involving base contacts from a few α-helical positions. In the following paper, we describe a complementary technique which confirms the identity of amino acids capable of DNA sequence discrimination from these positions.

The manner in which DNA-binding protein domains are able to discriminate between different DNA sequences is an important question in understanding crucial processes such as the control of gene expression in differentiation and development. The zinc finger motif has been studied extensively, with a view to providing some insight into this problem, owing to its remarkable prevalence in the eukaryotic genome, and its important role in proteins which control gene expression in *Drosophila* (*eg* 1), the mouse (2) and humans (3).

Most sequence-specific DNA-binding proteins bind to the DNA double helix by inserting an α-helix into the major groove (4, 5, 6). Sequence specificity results from the geometrical and chemical complementarity between the amino acid side chains of the α-helix and the accessible groups exposed on the edges of base-pairs. In addition to this direct reading of the DNA sequence, interactions with the DNA backbone stabilise the complex and are sensitive to the conformation of the nucleic acid, which in turn depends on the base sequence (7). *A priori*, a simple set of rules might suffice to explain the specific association of protein and DNA in all complexes, based on the possibility that certain amino acid side chains have preferences for particular base-pairs. However, crystal structures of protein-DNA complexes have shown that proteins can be idiosyncratic in their mode of DNA recognition, because they use alternative geometries to present their sensory α-helices to DNA, allowing a variety of different base contacts to be made by a single amino acid and *vice versa* (8). Nevertheless, for a family of transcription factors which use a "probe helix" for binding to the major groove of DNA, it would seem possible to deduce some general principles (9).

We believe the zinc finger of the TFIIIA class to be a good candidate for deriving a set of specificity rules owing to its great simplicity of structure and interaction with DNA. The zinc finger is an independently folding domain which uses a zinc ion to stabilise the packing of an antiparallel β-sheet against an α-helix (10, 11, 12). The crystal structures of zinc finger-DNA complexes show a semiconserved pattern of interactions in which 3 amino acids from the α-helix contact 3 adjacent bases (a triplet) in DNA (13, 14, 15). Thus the mode of DNA recognition is principally a one to one interaction between

amino acids and bases. Because zinc fingers function as independent modules (10, 16), fingers with different triplet specificities are combined to give specific recognition of longer DNA sequences. Protein engineering experiments have shown that it is possible to alter rationally the DNA-binding characteristics of individual zinc fingers when one or more of the α-helical positions is varied in a number of proteins (17, 18, 19). Because a large collection of these mutants is accumulating, it has already been possible to propose some rules relating amino acids on the α-helix to corresponding bases in the bound DNA sequence (20). However in this approach the altered positions on the α-helix are prejudged, making it possible to overlook the role of positions which are not currently considered important; and secondly, owing to the importance of context, concomitant alterations are sometimes required to affect specificity (20), so that a significant correlation between an amino acid and base may be misconstrued.

An alternative to the rational but biased design of proteins with new specificities, is the isolation of desirable mutants from a large pool. A powerful method of selecting such proteins is the cloning of peptides (21), or protein domains (22, 23), as fusions to the minor coat protein (pIII) of bacteriophage fd, which leads to their expression on the tip of the capsid. Phage displaying the peptides of interest can then be affinity purified and amplified for use in further rounds of selection and for DNA sequencing of the cloned gene. We have applied this technology to the study of zinc finger-DNA interactions, after demonstrating that functional zinc finger proteins can be displayed on the surface of fd phage, and that the engineered phage can be captured on a solid support coated with specific DNA. A phage display library has been created comprising variants of the middle finger from the DNA binding domain of Zif268 (a mouse transcription factor containing 3 zinc fingers) (2). DNA of fixed sequence is used to purify phage from this library over several rounds of selection, returning a number of different but related zinc fingers which bind the given DNA. By comparing similarities in the amino acid sequences of functionally equivalent fingers we deduce the likely mode of interaction of these fingers with DNA. Remarkably, it would appear that many base contacts can occur from three primary positions on the α-helix of the zinc finger, correlating with the implications of

the crystal structure of Zif268 bound to DNA (13). The ability to select or design zinc fingers with desired specificity means that in the near future, DNA binding proteins containing zinc fingers will be made to measure.

MATERIALS AND METHODS

Construction and cloning of genes. The gene for the first three fingers (residues 3-101) of Transcription Factor IIIA (TFIIIA) was amplified by PCR from the cDNA clone of TFIIIA using forward and backward primers which contain restriction sites for *Not*I and *Sfi*I respectively. The gene for the Zif268 fingers (residues 333-420) was assembled from 8 overlapping synthetic oligonucleotides, giving *Sfi*I and *Not*I overhangs. The genes for fingers of the phage library were synthesised from 4 oligonucleotides by directional end to end ligation using 3 short complementary linkers, and amplified by PCR from the single strand using forward and backward primers which contained sites for *Not*I and *Sfi*I respectively. Backward PCR primers in addition introduced Met-Ala-Glu as the first three amino acids of the zinc finger peptides, and these were followed by the residues of the wild type or library fingers as discussed in the text. Cloning overhangs were produced by digestion with *Sfi*I and *Not*I where necessary. Fragments were ligated to 1µg similarly prepared Fd-Tet-SN vector. This is a derivative of fd-tet-DOG1 (24) in which a section of the pelB leader and a restriction site for the enzyme *Sfi*I (underlined) have been added by site-directed mutagenesis using the oligonucleotide

5'CTCCTGCAGTTGGACCTGTGCC ATGGCCGGCTGGGCCGCATAGAATGGAACAACTAAAGG3'

which anneals in the region of the polylinker (L. Jespers, personal communication). Electrocompetent DH5α cells were transformed with recombinant vector in 200ng aliquots, grown for 1 hour in 2xTY medium with 1% glucose, and plated on TYE containing 15µg/ml tetracycline and 1% glucose.

Phage selection. Colonies were transferred from plates to 200ml 2xTY/Zn/Tet (2xTY containing 50µM $Zn(CH_3.COO)_2$ and 15µg/ml tetracycline) and grown overnight. Phage were purified from the culture supernatant by two rounds of precipitation using 0.2 volumes of 20% PEG/2.5M NaCl containing 50µM $Zn(CH_3.COO)_2$, and resuspended in zinc finger phage buffer (20mM HEPES pH7.5, 50mM NaCl, 1mM $MgCl_2$ and 50µM $Zn(CH_3.COO)_2$). Streptavidin-coated paramagnetic beads (Dynal) were washed in zinc finger phage buffer and blocked for 1 hour at room temperature with the same buffer

made up to 6% in fat-free dried milk (Marvel). Selection of phage was over three rounds: in the first round, beads (1mg) were saturated with biotinylated oligonucleotide (~80nM) and then washed prior to phage binding, but in the second and third rounds 1.7nM oligonucleotide and 5µg poly dGC (Sigma) were added to the beads with the phage. Binding reactions (1.5ml) for 1 hour at 15°C were in zinc finger phage buffer made up to 2% in fat-free dried milk (Marvel) and 1% in Tween 20, and typically contained $5x10^{11}$ phage. Beads were washed 15 times with 1ml of the same buffer. Phage were eluted by shaking in 0.1M triethylamine for 5min and neutralised with an equal volume of 1M Tris pH7.4. Log phase *E. coli* TG1 in 2xTY were infected with eluted phage for 30min at 37°C and plated as described above. Phage yields were titred by plating serial dilutions of the infected bacteria.

Sequencing of selected phage. Single colonies of transformants obtained after three rounds of selection as described, were grown overnight in 2xTY/Zn/Tet. Small aliquots of the cultures were stored in 15% glycerol at -20°C, to be used as an archive. Single-stranded DNA was prepared from phage in the culture supernatant and sequenced using Sequenase 2.0 (U.S. Biochemical Corp.).

RESULTS AND DISCUSSION

Phage display of 3-finger DNA-Binding Domains from TFIIIA or Zif268. Prior to the construction of a phage display library, we demonstrated that peptides containing three fully functional zinc fingers could be displayed on the surface of viable fd phage when cloned in the vector Fd-Tet-SN. In preliminary experiments, we cloned as fusions to pIII firstly the three N-terminal fingers from TFIIIA (25), and secondly the three fingers from Zif268 (2), for both of which the DNA binding sites are known. Peptide fused to the minor coat protein was detected in Western blots using an anti-pIII antibody (26). Approximately 10-20% of total pIII in phage preparations was present as fusion protein.

Phage displaying either set of fingers were capable of binding to specific DNA oligonucleotides, indicating that zinc fingers were expressed and correctly folded in both instances. Paramagnetic beads coated with specific oligonucleotide were used as a medium on which to capture DNA-binding phage (Fig.1A & C), and were consistently able to return between 100 and 500-fold more such phage, compared to free beads or beads coated with non-specific DNA. Alternatively, when phage displaying the three fingers of Zif268 were diluted $1:1.7 \times 10^3$ with Fd-Tet-SN phage not bearing zinc fingers, and the mixture incubated with beads coated with Zif268 operator DNA, one in three of the total phage eluted and transfected into *E. coli* were shown by colony hybridisation to carry the Zif268 gene, indicating an enrichment factor of over 500 for the zinc finger phage. Hence it is clear that zinc fingers displayed on fd phage are capable of preferential binding to DNA sequences with which they can form specific complexes, making possible the enrichment of wanted phage by factors of up to 500 in a single affinity purification step. Therefore, over multiple rounds of selection and amplification, very rare clones capable of sequence-specific DNA binding can be selected from a large library.

A phage display library of zinc fingers from Zif268. We have made a phage display library of the three fingers of Zif268 in which selected residues in the middle finger are randomised (Fig. 1B), and have isolated phage bearing zinc fingers with

desired specificity using a modified Zif268 operator sequence (27) in which the middle DNA triplet is altered to the sequence of interest (Fig.1C). In order to be able to study both the primary and secondary putative base recognition positions which are suggested by database analysis (28), we have designed the library of the middle finger so that, relative to the first residue in the α-helix (position +1), positions -1 to +8, but excluding the conserved Leu and His, can be any amino acid except Phe, Tyr, Trp and Cys which occur rarely at those positions (29). In addition, we have allowed position +9 (which might make an inter-finger contact with Ser at position -2 (13)) to be either Arg or Lys, the two most frequently occurring residues at that position.

The logic of this protocol, based upon the Zif268 crystal structure (13), is that the randomised finger is directed to the central triplet since the overall register of protein-DNA contacts is fixed by its two neighbours. This enables us to examine which amino acids in the randomised finger are the most important in forming specific complexes with DNA of known sequence. Since comprehensive variations are programmed in all the putative contact positions of the α-helix, we are able to conduct an objective study of the importance of each position in DNA-binding (28).

The size of the phage display library required assuming full degeneracy of the 8 variable positions is ($16^7 \times 2^1 =$) 5.4×10^8, but because of practical limitations in the efficiency of transformation with Fd-Tet-SN, we have been able to clone only 2.6×10^6 of these. The library we use is therefore some two hundred times smaller than the theoretical size necessary to cover all the possible variations of the α-helix. Despite this shortfall, it has been possible to isolate phage which bind with high affinity and specificity to given DNA sequences, demonstrating the remarkable versatility of the zinc finger motif.

Amino acid-base contacts in zinc finger-DNA complexes deduced from phage display selection. Of the 64 base triplets that could possibly form the binding site for variations of finger 2, we have so far used 32 in attempts to isolate zinc finger phage as described. Results from these selections are shown in Table 1. In general we have been unable to select zinc fingers which bind specifically to triplets without a 5'

or 3' guanine, all of which return the same limited set of phage after three rounds of selection (see legend to Table 1). However for each of the other triplets used to screen the library, a family of zinc finger phage is recovered. In these families, we find a sequence bias in the randomised α-helix, which we interpret as revealing the position and identity of amino acids used to contact the DNA. For instance: the middle fingers from the 8 different clones selected with the triplet GAT (Table1*d*) all have Asn at position +3 and Arg at position +6, just as does the first zinc finger of the *Drosophila* protein *tramtrack* in which they are seen making contacts to the same triplet in the cocrystal with specific DNA (14). This indicates that the positional recurrence of a particular amino acid in functionally equivalent fingers is unlikely to be coincidental, but rather because it has a functional role. Thus using data collected from the phage display library (Table 1) it is possible to infer most of the specific amino acid-DNA interactions. Remarkably, most of the results can be rationalised in terms of contacts from the three primary α-helical positions (-1, +3 and +6) identified by X-ray crystallography (13) and database analysis (28).

As has been pointed out before (30), guanine has a particularly important role in zinc finger-DNA interactions. When present at the 5' (*e.g.* Table1*c-i*) or 3' (*e.g.* Table1*m-o*) end of a triplet, G selects fingers with Arg at position +6 or -1 of the α-helix respectively. When present in the middle position of a triplet (*e.g.* Table1*b*), G prefers His at position +3. Occasionally, G at the 5' end of a triplet selects Ser or Thr at +6 (*e.g.* Table1*p*). Since G can only be specified absolutely by Arg (31), this is the most common determinant at -1 and +6. We can expect this type of contact to be a bidentate hydrogen bonding interaction as seen in the crystal structures of Zif268 (13) and *tramtrack* (14). In these structures, and in almost all of the selected fingers in which Arg recognises G at the 3' end, Asp occurs at position +2 to buttress the long Arg side chain (*e.g.* Table1*o,p*). When position -1 is not Arg, Asp rarely occurs at +2, suggesting that in this case any other contacts it might make with the second DNA strand do not contribute significantly to the stability the protein-DNA complex.

Adenine is also an important determinant of sequence specificity, recognised almost exclusively by Asn or Gln which again are able to make bidentate contacts (31). When A is present at the 3' end of a triplet, Gln is often selected at position -1 of the α-helix, accompanied by small aliphatic residues at +2 (*e.g.* Table1*b*). Adenine in the middle of the triplet strongly selects Asn at +3 (*e.g.* Table1*c-e*), except in the triplet CAG (Table1*a*) which selected only two types of finger, both with His at +3 (one being the wild-type Zif268 which contaminated the library during this experiment). The triplets ACG (Table1*j*) and ATG (Table1*k*), which have A at the 5' end, also returned oligoclonal mixtures of phage, the majority of which were of one clone with Asn at +6.

In theory, cytosine and thymine cannot be reliably discriminated by a hydrogen bonding amino acid side chain in the major groove (31). Nevertheless, C in the 3' position of a triplet shows a marked preference for Asp or Glu at position -1, together with Arg at +1 (*e.g.* Table1*e-g*). Asp is also sometimes selected at +3 and +6 when C is in the middle (*e.g.* Table1*o*) and 5' (*e.g.* Table1*a*) position respectively. Although Asp can accept a hydrogen bond from the amino group of C, we should note that the positive molecular charge of C in the major groove (32) will favour an interaction with Asp regardless of hydrogen bonding contacts. However, C in the middle position most frequently selects Thr (*e.g.* Table1*i*), Val or Leu (*e.g.* Table1*o*) at +3. Similarly, T in the middle position most often selects Ser (*e.g.* Table1*l*), Ala or Val (*e.g.* Table1*p*) at +3. The aliphatic amino acids are unable to make hydrogen bonds but Ala probably has a hydrophobic interaction with the methyl group of T, whereas a longer side chain such as Leu can exclude T and pack against the ring of C. When T is at the 5' end of a triplet, Ser and Thr are selected at +6 (as is occasionally the case for G at the 5' end). Thymine at the 3' end of a triplet selects a variety of polar amino acids at -1 (*e.g.* Table1*d*), and occasionally returns fingers with Ser at +2 (*e.g.* Table1*d*) which could make a contact as seen in the *tramtrack* crystal structure (14).

Limitations of phage display. From Table 1 it can be seen that a consensus or bias usually occurs in two of the three primary positions (-1, +3 and +6) for any family of equivalent fingers, suggesting that in many cases phage selection is by virtue of

only two base contacts per finger as is observed in the Zif268 crystal structure (13). Accordingly, identical finger sequences are often returned by DNA sequences differing by one base in the central triplet. One reason for this is that the phage display selection, being essentially purification by affinity, can yield zinc fingers which bind equally tightly to a number of DNA triplets and so are unable to discriminate. Secondly, since complex formation is governed by the law of mass action, affinity selection can favour those clones whose representation in the library is greatest even though their true affinity for DNA is less than that of other clones less abundant in the library. Phage display selection by affinity is therefore of limited value in distinguishing between permissive and specific interactions beyond those base contacts necessary to stabilise the complex. Thus in the absence of competition from fingers which are able to bind specifically to a given DNA, the tightest non-specific complexes will be selected from the phage library. Consequently, results obtained by phage display selection from a library must be confirmed by specificity assays, particularly when that library is of limited size.

Conclusion. The amino acid sequence biases observed within a family of functionally equivalent zinc fingers indicate that, of those α-helical positions randomised in this study, only three primary (-1, +3 and +6) and one auxiliary (+2) positions are involved in the recognition of DNA. Moreover, a limited set of amino acids are to be found at those positions, and we presume that these make contacts to bases. The indications therefore are that a code can be derived to describe zinc finger-DNA interactions. At this stage however, although sequence homologies are strongly suggestive of amino acid preferences for particular base-pairs, we cannot confidently deduce such rules until the specificity of individual fingers for DNA triplets is confirmed. We therefore defer making a summary table of these preferences until the following paper (33) in which we describe how randomised DNA binding sites can be used to this end.

While this work was in progress, a paper appeared by Rebar and Pabo (34) in which phage display was also used to select zinc fingers with new DNA-binding specificities. These authors constructed a library in which the first finger of Zif268 is

randomised, and screened with tetranucleotides to take into account end effects such as additional contacts from variants of this finger. Only 4 positions (-1, +2, +3 and +6) were randomised, chosen on the basis of the earlier X-ray crystal structures. The results of our work in which more positions were randomised, to some extent justifies Rebar and Pabo's use of the four random positions without apparent loss of effect, although further selections may reveal that the library is compromised. However, randomising only four positions decreases the theoretical library size so that full degeneracy can be achieved in practice. Nevertheless we find that the results obtained by Rebar and Pabo by screening their complete library with two variant Zif268 operators, are in agreement with our conclusions derived from an incomplete library. On the one hand this again highlights the versatility of zinc fingers but, remarkably, so far both studies have been unable to produce fingers which bind to the sequence CCT. It will be interesting to see whether sequence biases such as we have detected would be revealed, if more selections were performed using Rebar and Pabo's library. In any case, it would be desirable to investigate the effects on selections of using different numbers of randomised positions in more complete libraries than we have used at present.

REFERENCES FOR SECTION 1

1. Harrison, S. D. & Travers, A. A. (1990) *EMBO J.* **9**, 207-216.
2. Christy, B. A., Lau, L. F. & Nathans, D. (1988) *Proc. Natl. Acad. Sci.USA* **85**, 7857-7861.
3. Kinzler, K. W., Ruppert, J. M., Bigner, S. H. & Vogelstein, B. (1988) *Nature (London)* **332**, 371.
4. Pabo, C. O. & Sauer, R. T. (1992) *Annu.Rev.Biochem.* **61**, 1053-1095.
5. Harrison, S. C. (1991) *Nature (London)* **353**, 715-719.
6. Klug, A. (1993) *Gene* **135**, 83-92.
7. Dickerson, R. E. & Drew, H. R. (1981) *J. Mol. Biol.* **149**, 761-786.
8. Matthews, B. W. (1988) *Nature (London)* **335**, 294-295.
9. Suzuki, M. (1994) *Structure* **2**, 317-326.
10. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609-1614.
11. Berg, J. M. (1988) *Proc. Natl. Acad. Sci.USA* **85**, 99-102.
12. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245**, 635-637.
13. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809-817.
14. Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T. & Rhodes, D. (1993) *Nature (London)* **366**, 483-487.
15. Pavletich, N. P. & Pabo, C. O. (1993) *Science* **261**, 1701-1707.
16. Klug, A. & Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464-469.
17. Nardelli, J., Gibson, T. J., Vesque, C. & Charnay, P. (1991) *Nature (London)* **349**, 175-178.
18. Nardelli, J., Gibson, T. & Charnay, P. (1992) *Nucleic Acids Res.* **20**, 4137-4144.
19. Desjarlais, J. & Berg, J. (1992) *Proteins* **13**, 272.
20. Desjarlais, J. R. & Berg, J. M. (1992) *Proc. Natl. Acad. Sci.USA* **89**, 7345-7349.

21. Smith, G. P. (1985) *Science* 228, 1315-1317.

22. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990) *Nature (London)* 348, 552-554.

23. Bass, S., Greene, R. & Wells, J. A. (1990) *Proteins* 8, 309-314.

24. Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P. & Winter, G. (1991) *Nucleic Acids Res.*19, 4133-4137.

25. Ginsberg, A. M., King, B. O. & Roeder, R. G. (1984) *Cell* 39, 479-489.

26. Stengele, I., Bross, P., Garces, X., Giray, J. & Rasched, I. (1990) *J. Mol. Biol.* 212, 143-149.

27. Christy, B. & Nathans, D. (1989) *Proc. Natl. Acad. Sci.USA* 86, 8737-8741.

28. Jacobs, G. H. (1992) *EMBO J.* 11, 4507-4517.

29. Jacobs, G. (1993) *Ph.D. thesis, Cambridge*

30. Berg, J. M. (1992) *Proc. Natl. Acad. Sci.USA* 89, 11109-11110.

31. Seeman, N. D., Rosenberg, J. M. & Rich, A. (1976) *Proc. Nat. Acad. Sci.USA* 73, 804-808.

32. Hunter, C. A. (1993) *J. Mol. Biol.* 230, 1025-1054.

33. Choo, Y. & Klug, A. (1994) *Proc Natl.Acad.Sci.USA* (following paper - Section 2)

34. Rebar, E. J. & Pabo, C. O. (1994) *Science* 263, 671-673.

35. Marks, J. D., Hoogenboom. H. R., Griffiths. A. D. & Winter, G. (1992) *J. Biol. Chem.*267,16007-10.

FIGURE LEGENDS FOR SECTION 1

FIG. 1 Affinity purification of zinc finger phage.

(A) Zinc fingers [A] are expressed on the surface of fd phage [B] as fusions to the minor coat protein [C]. Zinc finger phage are bound to 5'-biotinylated DNA oligonucleotide [D] attached to streptavidin-coated paramagnetic beads [E], and captured using a magnet [F]. (Figure adapted from Dynal AS and also Marks *et al.* (35))

(B) Protein sequence of the three zinc fingers from Zif268 used in the phage display library. The randomised positions in the α-helix of the second finger have residues marked 'X'. The amino acid positions are numbered relative to the first helical residue (position 1). For amino acids at positions -1 to +8, excluding the conserved Leu and His, codons are equal mixtures of (G,A,C)NN - T in first base position is omitted in order to avoid stop codons, but this has the effect that the codons for Trp, Phe, Tyr and Cys are not represented. Position +9 is specified by the codon A(G,A)G, allowing either Arg or Lys. Residues of the hydrophobic core are circled, whereas the zinc ligands are written as white letters on black circles. The positions forming the β-sheets and the α-helix of a zinc finger are marked below the sequence.

(C) Sequences of DNA oligonucleotides used to purify (i) phage displaying the first three fingers of TFIIIA, (ii) phage displaying the three fingers of Zif268, and (iii) zinc finger phage from the phage display library. The Zif268 consensus operator sequence used in the X-ray crystal structure (13) is highlighted in (ii), and in (iii) where 'X' denotes a base change from the ideal operator in oligonucleotides used to purify phage with new specificities. Biotinylation of one strand is shown by a circled 'B'.

Table 1. Amino acid sequences of the variant α-helical regions from clones of library phage selected after 3 rounds using variants of the Zif268 operator. The amino acid sequences, aligned in the one letter code, are listed alongside the DNA oligonucleotides used in their purification (*a-p*). The latter are denoted by the sequence of the central DNA triplet in the 'bound' strand of the variant Zif268 operator. The amino acid positions are numbered relative to the first helical residue (position 1), and the three primary

recognition positions are highlighted. The accompanying numbers indicate the independent occurrences of that clone in the sequenced population (5-10 colonies); where numbers are in parentheses, the clone(s) were detected in the penultimate round of selection but not in the final round. In addition to the DNA triplets shown here, others were also used in attempts to select zinc finger phage from the library, but most selected two clones, one having the α-helical sequence KASNLVSHIR, and the other having LRHNLETHMR. Those triplets were: ACT, AAA, TTT, CCT, CTT, TTC, AGT, CGA, CAT, AGA, AGC and AAT.

TABLE 1

			-1123456789
a	CAG	1	RGDHLKQHIK
		9	RSDHLTTHIR
b	TGA	3	QLAHLSTHKR
		1	QSVHLQSHSR
		(3)	QKGHLTEHRK
c	GAA	2	QGGNLVRHLR
		1	NGGNLGRHMK
		1	ARSNLLRHTR
		2	LQSNLVRHQR
		1	IASNLLRHQR
d	GAT	1	DRSNLERHTR
		1	NQSNLERHHR
		1	QQSNLVRHQR
		1	NGGNLGRHMK
		1	NGANLERHRR
		1	SQGNLQRHGR
		1	SHPNLNRHLK
		1	TPGNLTRHGR
e	GAC	4	DRSNLERHTR
		1	DHANLARHTR
f	GCC	2	DRSSLTRHTR
		7	ERGTLARHEK
		1	DRRLLQRHQR
g	GTC	6	DRSSLTRHTR
		1	ERTSLSRHIR
h	GCA	1	SAGTLVRHSK
		2	QAQTLQRHLK
		2	EKATLARHMK
		1	TGGSLARHER
i	GCT	1	RQSTLGRHTR
		1	EKATLARHMK
		1	QAQTLQRHLK
		1	ERGTLARHEK
		1	GRDALARHQK
		1	RGPDLARHGR
		1	SRDVLRRHNR

			-1123456789
j	ACG	8	RRDVLMNHIR
		1	RKDVLVSHVR
k	ATG	8	RRDVLMNHIR
		1	RGDALTSHER
		1	RVDALEAHRR
l	GTA	1	DRSSLTRHTR
		1	ERTSLSRHIR
		(1)	QARSLTRHQR
		(2)	TGGSLARHER
		2	QRASLASHMR
		1	NRDTLTRHSK
		(1)	ERGTLARHER
m	TTG	9	RGDALTSHER
		1	RADALMVHKR
n	CCG	5	RQDTLVGHER
		1	RQSTLVRHTR
		2	RAADLNRHVR
		1	RKDVLVSHVR
		1	RRDVLMNHIR
o	GCG	1	RSDTLKKHGK
		3	RGPDLARHGR
		1	AREVLQRHTR
		3	REDVLIRHGK
		1	RSDLLQRHHK
p	GTG	1	RLDGLRTHLK
		1	RGDALTSHER
		1	RADALMVHKR
		1	RVDALEAHRR
		1	RRDVLLNHIR
		2	REDVLIRHGK
		1	RSDLLQRHHK

SECTION 2

ABSTRACT In the preceding section we showed how selections from a library of zinc fingers displayed on phage yielded fingers able to bind to a number of DNA triplets. Here, we describe a new technique to deal efficiently with the converse problem, namely the selection of a DNA binding site for a given zinc finger. This is done by screening against libraries of DNA triplet binding sites randomised in two positions but having one base fixed in the third position. The technique is applied here to determine the specificity of fingers previously selected by phage display. We find that some of these fingers are able to specify a unique base in each position of the cognate triplet. This is further illustrated by examples of fingers which can discriminate between closely related triplets as measured by their respective equilibrium dissociation constants. Comparing the amino acid sequences of fingers which specify a particular base in a triplet, we infer that in most instances, sequence specific binding of zinc fingers to DNA can be achieved using a small set of amino acid-base contacts amenable to a code.

In principle, rules governing protein-DNA interactions can be deduced from a large database of correlations between the amino acid sequences of the proteins and the nucleotide sequences of their optimal binding sites. To this end, we have shown in the preceding paper (1) that functionally equivalent zinc fingers which bind to a given DNA sequence can be selected from a phage display library. However, determination of the optimal binding site for these fingers is still required, as a safeguard against spurious selections. One can determine the optimal binding sites of these (and other) proteins, by selection from libraries of randomised DNA. This approach, the principle of which is essentially the converse of zinc finger phage display, would provide an equally informative database from which the same rules can be independently deduced. However, until now the favored method for binding site determination, involving iterative selection and amplification of target DNA followed by sequencing, is a laborious process not conveniently applicable to the analysis of a large database (2, 3).

We present here a convenient and rapid new method which can reveal the optimal binding site(s) of a DNA binding protein by single step selection from small libraries, and use this to check the binding site preferences of those zinc fingers selected previously by phage display (1). For this application, we use 12 different mini-libraries of the Zif268 binding site, each one with the central triplet having one position defined with a particular base pair and the other two positions randomised. Each library therefore comprises 16 oligonucleotides and offers a number of potential binding sites to the middle finger, provided that the latter can tolerate the defined base pair. Each zinc finger phage is screened against all 12 libraries individually immobilised in wells of a microtitre plate, and binding is detected by an enzyme immunoassay. Thus a pattern of acceptable bases at each position is disclosed, which we call a 'binding site signature'. The information contained in a binding site signature encompasses the repertoire of binding sites recognised by a zinc finger.

The binding site signatures obtained, using zinc finger phage selected as described in the preceding paper (1), reveal that the selection has yielded some highly sequence-specific zinc fingers which discriminate at all three positions of a triplet. From

measurements of equilibrium dissociation constants we find that these fingers bind tightly to the triplets indicated in their signatures, and discriminate against closely related sites usually by at least a factor of ten. The binding site signatures allow us to infer rules towards a specificity code for the interactions of zinc fingers with DNA.

MATERIALS AND METHODS

Binding site signatures. Flexible flat-bottomed 96-well microtitre plates (Falcon) were coated overnight at 4°C with streptavidin (0.1mg/ml in 0.1M $NaHCO_3$ pH8.6, 0.03% NaN_3). Wells were blocked for one hour with PBS/Zn (PBS, 50µM $Zn(CH_3.COO)_2$) containing 2% fat-free dried milk (Marvel), washed 3 times with PBS/Zn containing 0.1% Tween, and another 3 times with PBS/Zn. The 'bound' strand of each oligonucleotide library was made synthetically and the other strand extended from a 5'-biotinylated universal primer using DNA polymerase I (Klenow fragment). Fill-in reactions were added to wells (0.8 pmole DNA library in each) in PBS/Zn for 15 minutes, then washed once with PBS/Zn containing 0.1% Tween, and once again with PBS/Zn. Overnight bacterial cultures each containing a selected zinc finger phage (1) were grown in 2xTY containing 50mM $Zn(CH_3.COO)_2$ and 15µg/ml tetracycline at 30°C. Culture supernatants containing phage were diluted tenfold by the addition of PBS/Zn containing 2% fat-free dried milk (Marvel), 1% Tween and 20 µg/ml sonicated salmon sperm DNA. Diluted phage solutions (50µl) were applied to wells and binding allowed to proceed for one hour at 20°C. Unbound phage were removed by washing 5 times with PBS/Zn containing 1% Tween, and then 3 times with PBS/Zn. Bound phage were detected as described (4), or using HRP-conjugated anti-M13 IgG (Pharmacia), and quantitated using SOFTmax 2.32 (Molecular Devices Corp.).

Determination of apparent equilibrium dissociation constants. Overnight bacterial cultures were grown in 2xTY/Zn/Tet at 30°C. Culture supernatants containing phage were diluted twofold by the addition of PBS/Zn containing 4% fat-free dried milk (Marvel), 2% Tween and 40 µg/ml sonicated salmon sperm DNA. Binding reactions, containing appropriate concentrations of specific 5'-biotinylated DNA and equal volumes of zinc finger phage solution, were allowed to equilibrate for 1h at 20°C. All DNA was captured on streptavidin-coated paramagnetic beads (500µg per well), which were subsequently washed 6 times with PBS/Zn containing 1% Tween, and then 3 times with PBS/Zn. Bound phage were detected using HRP-conjugated anti-M13 IgG

(Pharmacia) and developed as described (4). Optical densities were quantitated using SOFTmax 2.32 (Molecular Devices Corp.).

Estimations of the K_d are by fitting to the equation Kd=[DNA].[P]/[DNA.P], using the programme KaleidaGraph Version 2.0 (Abelbeck Software). Owing to the sensitivity of the ELISA used to detect protein-DNA complex, we can use zinc finger phage concentrations far below those of the DNA, as is required for accurate calculations of the K_d. The technique we use has the advantage that while the concentration of DNA (variable) must be known accurately, that of the zinc fingers (constant) need not be known (5). This circumvents the problem of calculating the number of zinc finger peptides expressed on the tip of each phage, although since only 10-20% of the gene III protein (pIII) carries such peptides we would expect on average less than one copy per phage. Binding is performed in solution to prevent any effects caused by the avidity (6) of phage for DNA immobilised on a surface. Moreover, in this case measurements of K_d by ELISA are made possible since equilibrium is reached in solution prior to capture on the solid phase.

RESULTS AND DISCUSSION

The binding site signature of the second finger of Zif268. The top row of Fig. 2 shows the signature of the second finger of wild type Zif268. From the pattern of strong signals indicating binding to oligonucleotide libraries having GNN, TNN, NGN and NNG as the middle triplet, it emerges that the optimal binding site for this finger is $T/_{G}GG$, in accord with the published consensus sequence (7). This has implications for the interpretation of the X-ray crystal structure of Zif268 solved in complex with a consensus operator having TGG as the middle triplet (8). For instance, His at position +3 of the middle finger was modelled as donating a hydrogen bond to N7 of G, suggesting an equivalent contact to be possible with N7 of A, but from the binding site signature we can see that there is discrimination against A. This implies that the His may prefer to make a hydrogen bond to O6 of G or a bifurcated hydrogen bond to both O6 and N7, or that a steric clash with the amino group of A may prevent a tight interaction with this base. Thus by considering the stereochemistry of double helical DNA, binding site signatures can give insight into the details of zinc finger-DNA interactions.

Amino acid-base contacts in zinc finger-DNA complexes deduced from binding site signatures. The binding site signatures of other zinc fingers (Fig.2) reveal that the phage selections we performed in our previous study (1) have yielded highly sequence specific DNA binding proteins. Some of these are able to specify a unique sequence for the middle triplet of a variant Zif268 binding site, and are therefore more specific than is Zif268 itself for its consensus site. Moreover, one can identify the fingers which recognise a particular oligonucleotide library, that is to say a specific base at a defined position, by looking down the columns of Fig. 2. By comparing the amino acid sequences of these fingers we can identify any residues which have genuine preferences for particular bases on bound DNA. With a few exceptions, these are as previously predicted on the basis of phage display (1), and are summarised in Table 2.

The binding site signatures also reveal an important feature of our phage display library which is crucial to the interpretation of our selection results. All the fingers in

our panel, regardless of the amino acid present at position +6, are able to recognise G or both G and T at the 5' end of a triplet. Our explanation for this is that the 5' position of the middle triplet is fixed as either G or T by a contact from the invariant Asp at position +2 of finger 3 to the partner of either base on the complementary strand, analogous to those seen in the Zif268 (8) and *tramtrack* (9) crystal structures (a contact to the NH_2 of C or A respectively in the major groove). Therefore Asp at position +2 of finger 3 is dominant over the amino acid present at position +6 of the middle finger, precluding the possibility of recognition of A or C at the 5' position. Future libraries must be designed with this interaction omitted or the position varied. Interestingly, given the framework of the conserved regions of the three fingers, we can identify a rule in the second finger which specifies a frequent interaction with both G and T, viz. the occurrence of Ser or Thr at position +6, which may donate a hydrogen bond to either base.

Modulation of base recognition by auxiliary positions. As we have noted above, position +2 is able to specify the base directly 3' of the 'cognate triplet', and can thus work in conjunction with position +6 of the preceding finger. The binding site signatures, whilst pointing to amino acid-base contacts from the three primary positions, indicate that auxiliary positions can play other parts in base recognition. A clear case in point is Gln at position -1, which is specific for A at the 3' end of a triplet when position +2 is a small non polar amino acid such as Ala, though specific for T when polar residues such as Ser are at position +2. The strong correlation between Arg at position -1 and Asp at position +2, the basis of which is understood from the X-ray crystal structures of zinc fingers (8, 9), is another instance of interplay between these two positions. Thus the amino acid at position +2 is able to modulate or enhance the specificity of the amino acid at other positions.

At position +3, a different type of modulation is seen in the case of Thr and Val, which most often prefer C in the middle position of a triplet, but in some zinc fingers are able to recognise both C and T. This ambiguity occurs possibly as a result of different hydrophobic interactions involving the methyl groups of these residues, and here a

flexibility in the inclination of the finger rather than an effect from another position *per se* may be the cause of ambiguous reading.

Quantitative measurements of dissociation constants. The binding site signature of a zinc finger reveals its differential base preferences at a given concentration of DNA. As the concentration of DNA is altered, one can expect the binding site signature of any clone to change, being more distinctive at low [DNA], and becoming less so at higher [DNA] as the K_d of less favourable sites is approached and further bases become acceptable at each position of the triplet. Furthermore, because two base positions are randomly occupied in any one library of oligonucleotides, binding site signatures are not formally able to exclude the possibility of context dependence for some interactions. Therefore to supplement binding site signatures, which are essentially comparative, quantitative determinations of the equilibrium dissociation constant of each phage for different DNA binding sites are required. After phage display selection and binding site signatures, these are the third and definitive stage in assessing the specificity of zinc fingers.

Examples of such studies presented in Fig. 3 reveal that zinc finger phages bind the operators indicated in their binding site signatures with K_ds in the range of 10^{-8}-10^{-9} M, and can discriminate against closely related binding sites by factors greater than an order of magnitude. Indeed, Fig. 3 shows such differences in affinity for binding sites which differ in only one out of nine base pairs. Since the zinc fingers in our panel were selected from a library by non-competitive affinity purification, there is the possibility that fingers which are even more discriminatory can be isolated using a competitive selection process.

Measurements of dissociation constants allow different triplets to be ranked in order of preference according to the strength of binding. The examples here indicate that the contacts from either position -1 or +3 can contribute to discrimination. Also, the ambiguity in certain binding site signatures referred to above can be shown to have a basis in the equal affinity of certain fingers for closely related triplets. This is

demonstrated by the K_ds of the finger containing the amino acid sequence RGDALTSHER for the triplets TTG and GTG.

A code for zinc finger-DNA recognition. One would expect that the versatility of the zinc finger motif will have allowed evolution to develop various modes of binding to DNA (and even to RNA), which will be too diverse to fall under the scope of a single code. However, although a code may not apply to all zinc finger-DNA interactions, there is now convincing evidence that a code applies to a substantial subset. This code will fall short of being able to predict unfailingly the DNA binding site preference of any given zinc finger from its amino acid sequence, but may yet be sufficiently comprehensive to allow the design of zinc fingers with specificity for a given DNA sequence.

Using the selection methods of phage display (1) and of binding site signatures, we find that in the case of Zif268-like zinc fingers, *DNA recognition involves four fixed principal* (three primary and one auxiliary) *positions on the α-helix, from where a limited and specific set of amino acid-base contacts result in recognition of a variety of DNA triplets.* In other words, a code can describe the interactions of zinc fingers with DNA. Towards this code, we can propose amino acid-base contacts for almost all the entries in a matrix relating each base to each position of a triplet (Table 2). Where there is overlap, our results complement those of Desjarlais and Berg who have derived similar rules by altering zinc finger specificity using database-guided mutagenesis (10, 11).

Combinatorial use of the coded contacts. The individual base contacts listed in Table 2, though part of a code, may not always result in sequence specific binding to the expected base triplet when used in any combination. In the first instance we must be aware of the possibility that zinc fingers may not be able to recognise certain combinations of bases in some triplets by use of this code, or even at all. Otherwise, the majority of inconsistencies may be accounted for by considering variations in the inclination of the trident reading head of a zinc finger with respect to the triplet with which it is interacting. It appears that the identity of an amino acid at any one α-helical position is attuned to the identity of the residues at the other two positions to allow three

base contacts to occur simultaneously. Therefore, for example, in order that Ala may pick out T in the triplet GTG, Arg must not be used to recognise G from position +6, since this would distance the former too far from the DNA (see for example the finger containing the amino acid sequence RGDALTSHER). Secondly, since the pitch of the α-helix is 3.6 amino acids per turn, positions -1, +3 and +6 are not an integral number of turns apart, so that position +3 is nearer to the DNA than are -1 or +6. Hence, for example, short amino acids such as His and Asn, rather than the longer Arg and Gln, are used for the recognition of purines in the middle position of a triplet.

As a consequence of these distance effects we might say that the code is not really 'alphabetic' (always identical amino acid:base contact) but rather 'syllabic' (use of a small repertoire of amino acid:base contacts). An alphabetic code would involve only four rules, but syllabicity adds an additional level of complexity, since systematic combinations of rules comprise the code. Nevertheless, the recognition of each triplet is still best described by a code of syllables, rather than a catalogue of 'logograms' (idiosyncratic amino acid:base contact depending on triplet).

Conclusions. The 'syllabic' code of interactions with DNA is made possible by the versatile framework of the zinc finger: this allows an adaptability at the interface with DNA by slight changes of orientation, which in turn maintains a stoichiometry of one coplanar amino acid per base-pair in many different complexes. Given this mode of interaction between amino acids and bases it is to be expected that recognition of G and A by Arg and Asn/Gln respectively are important features of the code; but remarkably, other interactions can be more discriminatory than was anticipated (12). Conversely, it is clear that degeneracy can be programmed in the zinc fingers in varying degrees, allowing for intricate interactions with different regulatory DNA sequences (13, 7). One can see how this principle makes possible the regulation of differential gene expression by a limited set of transcription factors.

As we have noted, the versatility of the finger motif will likely allow other modes of binding to DNA. Similarly, we must take into account the malleability of nucleic acids, such as is observed in Fairall *et al.* (9) where a deformation of the double helix at a

flexible base step allows a direct contact from Ser at position +2 of finger1 to a T at the 3' position of the cognate triplet. Even in our selections there are instances of fingers whose binding mode is obscure, and may require structural analyses for clarification. Thus, water may be seen to play an important role, for example where short side chains such as Asp, Asn or Ser interact with bases from position -1 (14, 15).

Eventually, it might be possible to develop a number of codes describing zinc finger binding to DNA, which could predict the binding site preferences of some zinc fingers from their amino acid sequence. The functional amino acids selected at positions -1, +3 and to an extent +6 in this study, are very frequently observed at the same positions in naturally occurring fingers (*e.g.* see Fig. 4. in Desjarlais and Berg (16)), supporting the existence of coded contacts from these three positions. However, the lack of definitive predictive methods is not a serious practical limitation as current laboratory techniques (here and in (2, 3)) will allow the identification of binding sites for a given DNA-binding protein. Rather, we can apply phage selection and a knowledge of the recognition rules to the converse problem, namely the design of proteins to bind predetermined DNA sites.

Prospects for the design of DNA-binding proteins. The ability to manipulate the sequence specificity of zinc fingers implies that we are on the eve of designing DNA-binding proteins with desired specificity for applications in medicine and research (11, 17). This is possible because, by contrast to all other DNA-binding motifs, we can avail ourselves of the modular nature of the zinc finger, since DNA sites can be recognised by appropriate combinations of independently acting fingers linked in tandem.

The coded interactions of zinc fingers with DNA can be used to model the specificity of individual zinc fingers *de novo* , or more likely in conjunction with phage display selection of suitable candidates. In this way, according to requirements, one could modulate the affinity for a given binding site, or even engineer an appropriate degree of indiscrimination at particular base positions. Moreover, the additive effect of multiply repeated domains offers the opportunity to bind specifically and tightly to extended, and

hence very rare, genomic loci. Thus zinc finger proteins might well be a good alternative to the use of antisense nucleic acids in suppressing or modifying the action of a given gene, whether normal or mutant. To this end, extra functions could be introduced to these DNA binding domains by appending suitable natural or synthetic effectors.

REFERENCES FOR SECTION 2

1. Choo Y. & Klug, A. (1994) *Proc.Natl. Acad.Sci.USA* (preceding section)
2. Thiesen, H., J & Bach, C. (1990) *Nucleic Acids Res.*18, 3203-3209.
3. Pollock, R. & Treisman, R. (1990) *Nucleic Acids Res.* 18, 6197-6204.
4. Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I., Waterhouse, P., Crosby, W. L., Kontermann R. E., Jones, P. T., Low, N. M., Allison, T. J., Prospero, T. D., Hoogenboom, H. R., Nissim, A., Cox, J. P. L., Harrison, J. L., Zaccolo, M., Gherardi, E. & Winter, G. (1994) *EMBO J.* In press,
5. Choo, Y. & Klug, A. (1993) *Nucleic Acids Res.* 21, 3341-3346.
6. Marks, J. D., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1992) *J. Biol. Chem.* 267, 16007-16010.
7. Christy, B. & Nathans, D. (1989) *Proc. Natl. Acad. Sci.USA* 86, 8737-8741.
8. Pavletich, N. P. & Pabo, C. O. (1991) *Science* 252, 809-817.
9. Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T. & Rhodes, D. (1993) *Nature (London)* 366, 483-487.
10. Desjarlais, J. R. & Berg, J. M. (1992) *Proc. Natl. Acad. Sci. USA* 89, 7345-7349.
11. Desjarlais, J. R. & Berg, J. M. (1993) *Proc. Natl. Acad. Sci.USA* 90, 2256-2260.
12. Seeman, N. D., Rosenberg, J. M. & Rich, A. (1976) *Proc. Nat. Acad. Sci.USA* 73, 804-808.
13. Harrison, S. D. & Travers, A. A. (1990) *EMBO J.* 9, 207-216.
14. Qian, Y. Q., Otting, G. & Wuthrich, K. (1993) *J. Am. Chem. Soc.* 115, 1189-1190.
15. Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. & Sigler, P. B. (1994) *Nature (London)* 368, 469-473.
16. Desjarlais, J. R. & Berg, J. M. (1992) *Proteins* 12, 101-104.
17. Rebar, E. J. & Pabo, C. O. (1994) *Science* 263, 671-673.

FIGURE LEGENDS FOR SECTION 2

FIG 2 Binding site signatures of individual zinc finger phage. The diagram is of raw data and represents binding of zinc finger phage to randomised DNA immobilised in the wells of microtitre plates. To test each zinc finger phage against each oligonucleotide library (see text), DNA libraries are applied to columns of wells (down the plate), while rows of wells (across the plate) contain equal volumes of a solution of a zinc finger phage. The identity of each library is given as the middle triplet of the 'bound' strand of Zif268 operator, where N represents a mixture of all 4 nucleotides. The zinc finger phage is specified by the sequence of the variable region of the middle finger, numbered relative to the first helical residue (position 1), and the three primary recognition positions are highlighted. Bound phage are detected by an enzyme immunoassay. The approximate strength of binding is indicated by a grey scale proportional to the enzyme activity. From the pattern of binding to DNA libraries, called the 'signature' of each clone, one or a small number of binding sites can be read off and these are written on the right of the figure.

FIG 3. Determination of apparent equilibrium dissociation constants of zinc finger phage for variants of the Zif268 binding site, showing discrimination of closely related triplets by the middle finger, usually by factors of >10. The two outer fingers carry the native sequence, as do the two cognate outer DNA triplets. The sequence of amino acids occupying helical positions -1 to +9 of the varied middle finger is shown in each case.

Table 2. Summary of frequently observed amino acid-base contacts in interactions of selected zinc fingers with DNA. The given contacts comprise a syllabic recognition code (see text) for appropriate triplets. Cognate amino acids and their positions in the α-helix are entered in a matrix relating each base to each position of a triplet. Auxiliary amino acids from position +2 can enhance or modulate specificity of amino acids at position -1, and these are listed as pairs. Ser or Thr at position +6 permit Asp +2 *of the following finger* (denoted Asp ++2) to specify both G and T indirectly, and the pairs are listed. The

specificity of Ser+3 for T and Thr+3 for C may be interchangeable in rare instances, while Val+3 appears to be consistently ambiguous.

POSITION IN TRIPLET

NUCLEOTIDE	5'	Mid.	3'
G	Arg [Ser] [Thr]	His	Arg ♥
A		Asn	Gln ♠
T	[Ser] [Thr]	Ala {Ser} (Val)	Asn Gln ♣
C		Asp Leu {Thr} (Val)	Asp

TABLE 2

SECTION 3

Recently we have proposed that specific DNA-binding proteins comprising zinc fingers can be made to measure [1, 2]. To demonstrate their potential we have created a three finger peptide able to bind site-specifically to a unique 9bp region of a *BCR-ABL* fusion oncogene and to discriminate it from the parent genomic sequences [3]. Using transformed cells in culture as a model, we show that binding to the target oncogene in chromosomal DNA is possible, resulting in blockage of transcription. Consequently, murine cells made growth factor-independent by the action of the oncogene [4] are found to revert to factor dependence on transient transfection with a vector expressing the peptide.

DNA-binding proteins designed to recognise specific DNA sequences could be incorporated in chimeric transcription factors, recombinases, nucleases *etc.* for a wide range of applications. We have shown that zinc finger mini-domains can discriminate between closely related DNA triplets, and have proposed that they can be linked together to form domains for the specific recognition of longer DNA sequences [1, 2]. One interesting possibility for the use of such protein domains is to target selectively genetic differences in pathogens or transformed cells. Here we report one such application.

There exist a set of human leukaemias in which a reciprocal chromosomal translocation t(9;22) (q34;q11) results in a truncated chromosome 22, the Philadelphia chromosome (Ph^1) [5], encoding at the breakpoint a fusion of sequences from the c-*ABL* proto-oncogene [6] and the *BCR* gene [7]. In chronic myelogenous leukaemia (CML), the breakpoints usually occur in the first intron of the c-*ABL* gene and in the breakpoint cluster region of the *BCR* gene [8], and give rise to a $p210^{BCR\text{-}ABL}$ gene product [9]. Alternatively, in acute lymphoblastic leukaemia (ALL), the breakpoints usually occur in the first introns of both *BCR* and c-*ABL* [10], and result in a $p190^{BCR\text{-}ABL}$ gene product (Fig.4.) [11]. Facsimiles of these rearranged genes act as dominant transforming oncogenes in cell culture [4] and transgenic mice [12]. Like their genomic counterparts, the cDNAs bear a unique nucleotide sequence at the fusion point of the *BCR* and c-*ABL* genes, which can be recognised at the DNA level by a site-specific DNA-binding protein. We have designed such a protein to recognise the unique fusion site in the $p190^{BCR\text{-}ABL}$ cDNA. This fusion is obviously distinct from the breakpoints in the spontaneous genomic translocations, which are thought to be variable among patients. Although the design of such peptides has implications for cancer research, our primariy aim here is to prove the principle of protein design, and to assess the feasibility of *in vivo* binding to chromosomal DNA in available model systems.

The DNA-binding proteins we create are composed of classical zinc fingers [13, 14, 15]. These small motifs are ideal natural building blocks for *de novo* protein design since they function as independent modules [16], but can be connected by a well known linker [17] to allow recognition of long, asymmetric DNA sequences. Lately it has been possible to isolate zinc fingers which bind to given DNA triplets by selection from phage display libraries of

randomised zinc fingers [1, 18, 19]. The specificity of selected fingers is checked by a second selection technique called the 'binding site signature' in which these fingers are used to screen libraries of randomised oligonucleotide binding sites, thus identifying fingers which can specify a unique base triplet [2]. From these and other studies [20], elements of a recognition code have emerged which relate the amino acid sequence of zinc fingers to their cognate triplet.

The strategy we use in creating DNA-binding proteins combines phage display selection and rational design based on the available recognition rules. A nine base-pair target sequence (GCA.GAA.GCC) for a three zinc finger peptide was chosen which spanned the fusion point of the p190*BCR-ABL* cDNA [10]. The three triplets forming this binding site were each used to screen a zinc finger phage library over three rounds as described [1]. The selected fingers were then analysed by binding site signatures to reveal their prefered triplet, and mutations to improve specificity were made to the finger selected for binding to GCA (Fig.5). A phage display mini- library of putative *BCR-ABL*-binding three-finger proteins was cloned in fd phage, comprising six possible combinations of the six selected or designed fingers (1A, 1B; 2A; 3A, 3B and 3C) linked in the appropriate order. The mini library was screened once with an oligonucleotide containing the 9 base-pair *BCR-ABL* target sequence, to select for tight binding clones over weak binders and background vector phage. Because the library was small, we did not include competitor DNA sequences for homologous regions of the genomic *BCR* and c-*ABL* genes, but instead checked the selected clones for their ability to discriminate. We found that although all the selected clones were able to bind the *BCR-ABL* target sequence and to discriminate between this and the genomic *BCR* sequence, only a subset could discriminate against the c-*ABL* sequence which, at the junction between intron 1 and exon 2, has an 8/9 base-pair homology to the *BCR-ABL* target sequence [10]. Sequencing of the discriminating clones revealed two types of selected peptide, one with the composition 1A-2A-3B and the other with 1B-2A-3B. Thus both peptides carried the third finger (3B) which was specifically designed against the triplet GCA but peptide 1A-2A-3B was able to bind to the *BCR-ABL* target sequence with higher affinity than was peptide 1B-2A-3B.

The peptide 1A-2A-3B, which we will henceforth refer to as the anti-BCR-ABL peptide, was used in further experiments. The anti-BCR-ABL peptide has an apparent equilibrium dissociation constant (K_d) of 6.2±0.4 x 10^{-7}M for the p190$^{BCR-ABL}$ cDNA sequence *in vitro*, and discriminates against the similar sequences found in genomic *BCR* and c-*ABL* DNA, by factors greater than an order of magnitude (Fig. 6). The measured dissociation constant is higher than that of three-finger peptides from naturally occuring proteins such as Sp1 [21] or Zif268 [22], which have K_ds in the range of 10^{-9}M, but rather is comparable to that of the two fingers from the *tramtrack* (*ttk*) protein [23]. However, the affinity of the anti-BCR-ABL peptide could be refined, if desired, by site-directed mutations or by 'affinity maturation' of a phage display library [24].

Having established DNA discrimination *in vitro*, we wished to test whether the anti-BCR-ABL peptide was capable of site-specific DNA-binding *in vivo*. The peptide was fused to the VP16 activation domain from herpes simplex virus [25] and used in transient transfection assays (Fig.7) to drive production of a CAT (chloramphenicol acetyl transferase) reporter gene from a binding site upstream of the TATA box [26]. A thirty-fold increase in CAT activity was observed in cells cotransfected with reporter plasmid bearing copies of the p190$^{BCR-ABL}$ cDNA target site, compared to a barely detectable increase in cells cotransfected with reporter plasmid bearing copies of either the *BCR* or c-*ABL* semihomologous sequences. The selective stimulation of transcription indicates convincingly that highly site-specific DNA-binding can occur *in vivo*. However, while transient transfections assay binding to plasmid DNA, the true target site for this and most other DNA-binding proteins is in genomic DNA. This might well present significant problems, not least since this DNA is physically separated from the cytosol by the nuclear membrane, but also since it may be packaged within chromatin.

To study whether genomic targeting is possible, we made a construct in which our anti-BCR-ABL peptide was flanked at the N-terminus with the nuclear localisation signal from the large T antigen of SV40 virus [27], and at the C-terminus with an 11 amino acid c-myc epitope tag recognisable by the 9E10 antibody [28]. This construct was used to transiently transfect the IL-3-dependent murine cell line Ba/F3 [29], or alternatively Ba/F3+p190 and

Ba/F3+p210 cell lines previously made IL-3-independent by integrated plasmid constructs expressing either p190$^{BCR-ABL}$ or p210$^{BCR-ABL}$, respectively (ISG et al., in preparation) Staining of the cells with the 9E10 antibody followed by a secondary fluorescent conjugate showed efficient nuclear localisation in those cells transfected with the anti-*BCR-ABL* peptide (Fig.8). The efficiency of transient transfection, measured as the proportion of immunofluorescent cells in the population, was 15-20%. When IL-3 is withdrawn from tissue culture, a corresponding proportion of Ba/F3+p190 cells are found to have reverted to factor dependence and die, while Ba/F3+p210 cells are unaffected (Fig. 9a). Immunofluorescence microscopy of transfected Ba/F3+p190 cells in the absence of IL-3 shows chromatin condensation and nuclear fragmentation into small apoptotic bodies, while the nuclei of Ba/F3+p210 cells remain intact (Fig.8). Northern blots of total cytoplasmic RNA from Ba/F3+p190 cells transiently transfected with the anti-BCR-ABL peptide revealed reduced levels of p190$^{BCR-ABL}$ mRNA relative to untransfected cells. By contrast, similarly transfected Ba/F3+p210 cells showed no decrease in the levels of p210$^{BCR-ABL}$ mRNA (Fig. 9 b).

Hence a DNA-binding protein designed to recognise a specific DNA sequence *in vitro*, is active *in vivo* where, directed to the nucleus by an appended localisation signal, it can bind its target sequence in chromosomal DNA. This is found on otherwise actively transcribing DNA, so presumably binding of the peptide blocks the path of the polymerase, causing stalling or abortion. The use of a specific polypeptide in this case to target intragenic sequences is reminiscent of antisense oligonucleotide- or ribozyme- based approaches to inhibiting the expression of selected genes [30]. Like antisense oligonucleotides, zinc finger DNA-binding proteins can be tailored against genes altered by chromosomal translocations, or point mutations, as well as to regulatory sequences within genes. Also, like oligonucleotides which can be designed to repress transcription by triple helix formation in homopurine-homopyrimidine promoters [31], DNA-binding proteins can bind to various unique regions outside genes, but in contrast they can direct gene expression by both up- or down- regulating the initiation of transcription when fused to activation [32] or repression domains [33]. In any case, by acting directly on any DNA, and by allowing fusion to a variety

of protein effectors, tailored site-specific DNA-binding proteins have the potential to control gene expression, and indeed to manipulate the genetic material itself, in medicine and research.

FIGURE LEGENDS FOR SECTION 3

Fig. 4. Nucleotide sequences of the fusion point between *BCR* and *ABL* sequences in p190 cDNA, and of the corresponding exon boundaries in the *BCR* and c-*ABL* genes. Exon sequences are written in capital letters while introns are given in lowercase. Line1, p190$^{BCR-ABL}$ cDNA; line 2, *BCR* genomic sequence at junction of exon 1 and intron 1; line 3, *ABL* genomic sequence at junction of intron 1 and exon 2 [10]. The 9bp target sequence in the p190$^{BCR-ABL}$ cDNA is underlined, as are the homologous sequences in genomic *BCR* and c-*ABL*.

Fig. 5. Amino acid sequences of zinc fingers used in constructing the mini library of putative *BCR-ABL* binders. Regions of secondary structure are underlined below the list, while residue positions are given above, relative to the first position of the α-helix (position 1). Zinc finger phages were selected from a library of 2.6×10^6 variants, using three DNA binding sites each containing one of the triplets GCC, GAA or GCA [1]. Binding site signatures (data not shown) indicate that fingers 1A and 1B specify the triplet GCC, finger 2A specifies GAA, while the fingers selected using the triplet GCA all prefer binding to GCT [2]. Amongst the latter is finger 3A, the specificity of which we believed, on the basis of recognition rules, could be changed by a point mutation. Finger 3B, based on the selected finger 3A, but in which Gln at helical position +2 was altered to Ala should be specific for GCA. Finger 3C is an alternative version of finger 3A, in which the recognition of C is mediated by Asp+3 rather than by Thr+3.

Fig.6 . Discrimination in the binding of the anti-BCR-ABL peptide to its p190$^{BCR-ABL}$ target site and to like regions of genomic *BCR* and c-*ABL*. The graph shows binding (measured as an $A_{450-650}$) at various [DNA]. Binding reactions and complex detection by enzyme immunoassay were performed as described [2], and a full curve analysis was used in calculations of the K_d [17]. The DNAs used were oligonucleotides spanning 9bp either side of the fusion point in the cDNA or the exon boundaries. The anti-BCR-ABL peptide binds to its

intended target site with a $K_d = 6.2 \pm 0.4 \times 10^{-7}$M, and is able to discriminate against genomic *BCR* and c-*ABL* sequences, though the latter differs by only one base pair in the bound 9bp region.

Fig. 7. Transactivation reporter assays using an anti-BCR-ABL peptide fused to the VP16 activation domain. C3H10T$^1/_2$ cells were transiently cotransfected with a CAT reporter plasmid and an anti-BCR-ABL/VP16 expression vector (pZN1A). The top panel of the figure shows the results of thin layer chromatography of samples from different transfections, in which the fold induction of CAT activity relative to a sample where reporter alone was transfected (panel 1) is plotted on a histogram below. A specific (thirty-fold) increase in CAT activity was observed in cells cotransfected with reporter plasmid bearing copies of the p190$^{BCR\text{-}ABL}$ cDNA target site, indicating *in vivo* binding. The particular constructs used in different transfections are noted below the histogram.

METHODS: Reporter plasmids pMCAT6BA, pMCAT6A, and pMCAT6B, were constructed by inserting 6 copies of the p190$^{BCR\text{-}ABL}$ target site (CGCAGAAGCC), the c-*ABL* second exon-intron junction sequence (TCCAGAAGCC), or the *BCR* first exon-intron junction sequence (CGCAGGTGAG) respectively, into pMCAT3 [34]. The anti-BCR-ABL/VP16 expression vector was generated by inserting the in-frame fusion between the activation domain of herpes simplex virus VP16 [25] and the Zn finger peptide in the pEF-BOS vector [35]. C3H10T$^1/_2$ cells were transiently co-transfected with 10 µg of reporter plasmid and 10 µg of expression vector. RSVL [36], which contains the Rous sarcoma virus long terminal repeat linked to luciferase, was used as an internal control to normalise for differences in transfection efficiency. Cells were transfected by the calcium phosphate precipitation method and CAT assays performed as described [37]. Plasmid pG5EC, which has five consensus 17-mer GAL4-binding sites upstream from the minimal promoter of the adenovirus E1b TATA box, and pM1VP16 vector, which encodes an in-frame fusion between the DNA-binding domain of GAL4 and the activation domain of herpes simplex virus VP16, were used as a positive control [38].

Fig. 8. Immunofluorescence of Ba/F3+p190 and Ba/F3+p210 cells transiently transfected with the anti-bcr-abl expression vector and stained with the 9E10 antibody. The image shows expression and nuclear localisation of the anti-BCR-ABL peptide (panels B, C, and D). In addition, transfected Ba/F3+p190 cells show chromatin condensation and nuclear fragmentation into small apoptotic bodies (panels B, and C), but not either untransfected Ba/F3+p190 cells (panel A) or transfected Ba/F3+p210 cells (panel D).

METHODS: The anti-*BCR-ABL* expression vector was generated in the pEF-BOS vector [35], including an 11 amino acid c-myc epitope tag (EQKLISEEDLN) at the carboxy-terminal end, recognizable by the 9E10 antibody [28] and the nuclear localization signal PKKKRKV of the large T antigen of SV40 virus [27] at the amino-terminal end. Three glycine residues were introduced downstream of the nuclear localization signal as a spacer, to ensure exposure of the nuclear leader from the folded molecule. Ba/F3 cells were transfected with 25 μg of the anti-BCR-ABL expression construct tagged with the 9E10 c-myc epitope as described [39] and protein production analyzed 48 h later by immunofluorescence-labeling as follows. Cells were fixed in 4% (w/v) paraformaldehyde for 15 min, washed in phosphate-buffered saline (PBS), and permeabilized in methanol for 2 min. After blocking in 10% fetal calf serum in PBS for 30 min, the mouse 9E10 antibody was added. After a 30 min incubation at room temperature a fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse IgG (SIGMA) was added and incubated for a further 30 min. Fluorescent cells were visualized using a confocal scanning microscope (magnification, 200X).

Fig. 9a. Viability in the absence of IL-3 of transformed Ba/F3 cells transiently transfected with a vector expressing anti-*BCR-ABL* peptide. The Ba/F3 cell line is dependent on IL-3 for growth, but becomes IL-3 independent when stably transformed by p190$^{BCR-ABL}$ or p210$^{BCR-ABL}$ cDNA (4, and ISG et al., in preparation). A proportion of Ba/F3+p190 cells transfected with the anti-BCR-ABL expression vector revert to IL-3 dependence, while similarly transfected Ba/F3+p210 cells are unaffected.

METHODS: Cell lines Ba/F3, Ba/F3+p190 and Ba/F3+p210 were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum. In the case

of Ba/F3 cell line 10% WEHI-3B-conditioned medium was included as a source of IL-3. After the transfection with the anti-BCR-ABL expression vector, cells ($5x10^5$/ml) were washed twice in serum-free medium and cultured in DMEM medium with 10% fetal bovine serum without WEHI-3B-conditioned medium. Percentage viability was determined by tripan blue exclusion. Data are expressed as means of triplicate cultures.

Fig. 9b. Northern filter hybridisation analysis of Ba/F3+p190 and Ba/F3+p210 cell lines transfected with the anti-BCR-ABL expression vector. Lane 1 is from untransfected Ba/F3+p190 cell line; lanes 2, and 3 are from Ba/F3+p190 cell line transfected with the anti-BCR-ABL expression vector; lane 4 is from untransfected Ba/F3+p210 cell line; lanes 5, and 6 are from Ba/F3+p210 cell line transfected with the anti-BCR-ABL expression vector. When transfected with the anti-BCR-ABL expression vector, a specific downregulation of $p190^{BCR\text{-}ABL}$ mRNA is seen in Ba/F3+p190 cells, while expression of $p210^{BCR\text{-}ABL}$ is unaffected in Ba/F3+p210 cells.

METHODS: 10 μg of total cytoplasmic RNA, from the cells indicated, was glyoxylated and fractioned in 1.4% agarose gels in 10mM $NaPO_4$ buffer, pH 7.0. After electrophoresis the gel was blotted onto Hybond-N (Amersham), UV-cross linked and hybridized to an ^{32}P-labelled c-*ABL* probe. Autorradiography was for 14h at -70 ^{o}C. Loading was monitored by reprobing the filters with a mouse β-actin cDNA.

REFERENCES FOR SECTION 3

1. Choo, Y. & Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **in press**, (1994). (Section 1)
2. Choo, Y. & Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **in press**, (1994). (Section 2)
3. Kurzrock, R., Gutterman, J.U. & Talpaz, M. *N. Engl. J. Med.* **319**, 990-998 (1988).
4. Daley, G.Q. & Baltimore, D. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9312-9316 (1988).
5. Nowell, P.C. & Hungerford, D.A. *Science* **132**, 1197-1200 (1960).
6. Bartram, C.R., de Klein, A., Hagemeijer, A. & al, e. *Nature* **306**, 277-280 (1983).
7. Groffen, J., *et al. Cell* **36**, 93-99 (1984).
8. Shtivelman, E., Lifshitz, B., Gale, R.P. & Canaani, E. *Nature* **315**, 550-554 (1985).
9. Konopka, J.B., Watanabe, S.M. & Witte, O.N. *Cell* **37**, 1035-1042 (1984).
10. Hermans, A., Heisterkamp, N., von Linden & al, e. *Cell* **51**, 33-40 (1987).
11. Kurzrock, R., Shtalrid, M., Romero, R. & al, e. *Nature* **325**, 631-635 (1987).
12. Heisterkamp, N., *et al. Nature* **344**, 251-253 (1990).
13. Miller, J., McLachlan, A.D. & Klug, A. *EMBO J* **4**, 1609-1614 (1985).
14. Berg, J.M. *Proc. Natl. Acad. Sci.* **85**, 99-102 (1988).
15. Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. & Wright, P.E. *Science* **245**, 635-637 (1989).
16. Pavletich, N.P. & Pabo, C.O. *Science* **252**, 809-817 (1991).
17. Choo, Y. & Klug, A. *Nucleic Acids Research* **21**, 3341-46 (1993).
18. Rebar, E.J. & Pabo, C.O. *Science* **263**, 671-3 (1994).
19. Jamieson, A.C., Kim, S.-H. & Wells, J.A. *Biochemistry* **33**, 5689-5695 (1994).
20. Desjarlais, J.R. & Berg, J.M. *Proc. Natl. Acad. Sci. USA* **89**, 7345-7349 (1992).
21. Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. *Cell* **51**, 1079-1090 (1987).
22. Christy, B.A., Lau, L.F. & Nathans, D. *Proc. Natl. Acad. Sci.* **85**, 7857-7861 (1988).
23. Fairall, L., Harrison, S.D., Travers, A.A. & Rhodes, D. *J. Mol. Biol.* **226**, 349-366 (1992).
24. Hawkins, R.E., Russell, S.J. & Winter, G. *J. Mol. Biol.* **226**, 889-896 (1992).
25. Fields, S. *Methods* **5**, 116-124 (1993).
26. Gorman, C.M., Moffat, L.F. & Howard, B.H. *Mol. Cell. Biol.* **2**, 1044-1051 (1982).

27. Kalderon, D., Roberts, B.L., Richardson, W.D. & Smith, A.E. *Cell* 499-509 (1984).

28. Evan, G.I., Lewis, G.K., Ramsay, G. & Bishop, J.M. *Mol. Cell. Biol.* **5**, 3610-3616 (1985).

29. Palacios, R. & Steinmetz, M. *Cell* **41**, 727-734 (1985).

30. Stein, C.A. & Cheng, Y. *Science* **261**, 1004-1012 (1993).

31. Cooney, M., Czernuszewicz, G., Postel, E.H., Flint, S.J. & Hogan, M.E. *Science* **245**, 725-730 (1988).

32. Seipel, K., Georgiev, O. & Schaffner, W. *EMBO J.* **11**, 4961-4968 (1992).

33. Herschbach, B.M., Arnaud, M.B. & Johnson, A.D. *Nature* **370**, 309-311 (1994).

34. Luscher, B., Mitchell, P.J., Williams, T. & Tjian, R. *Genes Dev.* **3**, 1507-1517 (1989).

35. Mizushima, S. & Shigezaku, N. *Nucl. Acids Res.* **18**, 5322 (1990).

36. de Wet, J.R., Wood, K.V., de Luca, M., Helinski, D.R. & Subramani, S. *Mol. Cell Biol.* **7**, 725-737 (1987).

37. Sanchez-Garcia, I., Osada, H., Forster, A. & Rabbitts, T.H. *EMBO J.* **12**, 4243-4250 (1993).

38. Sadowski, I., Brendan, B., Broad, P. & Hollis, M. *Gene* **118**, 137-141 (1992).

39. Sanchez-Garcia, I. & Rabbitts, T.H. *Proc. Natl. Acad. Sci. U.S.A.* **in press**, (1994).

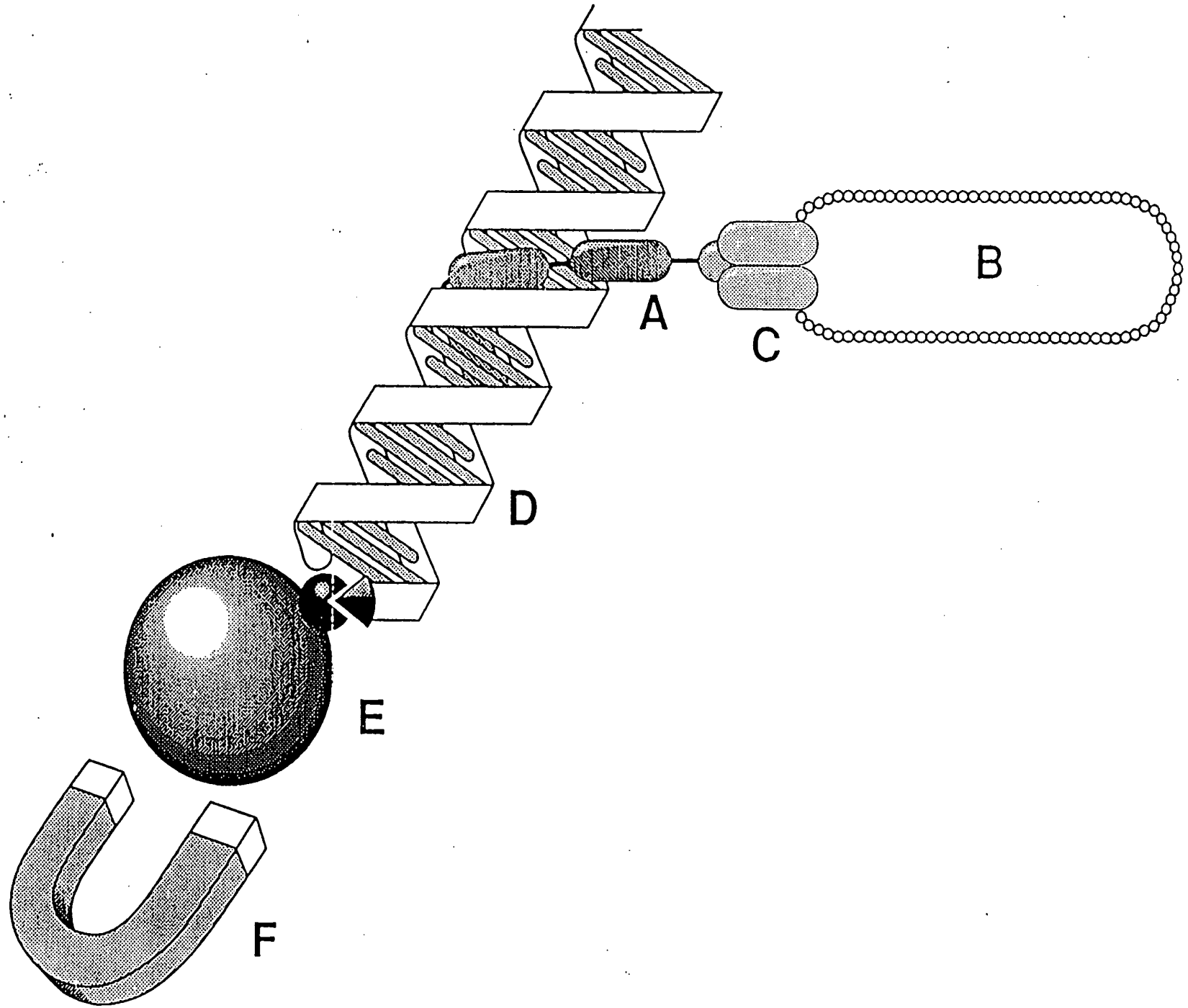


Fig. 1A

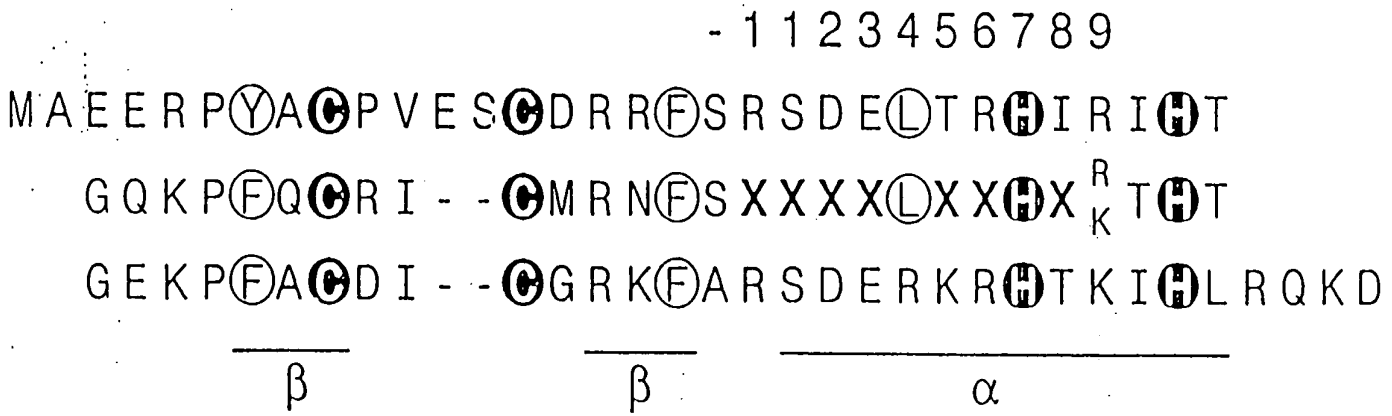

-1 1 2 3 4 5 6 7 8 9
MAEERPYACPVESCDRRFSRSDELTRHIRIHT
GQKPFQCRI--CMRNFSXXXXLXXHXR/KTHT
GEKPFACDI--CGRKFARSDERKRHTKIHLRQKD
β
β
α

Fig. 1B


(i)
TATGACTTGGATGGGAGACCGCCTGG
ACTGAACCTACCCTCTGGCGGACCTTAA
(ii)
TATATAGCGTGGGCGTATATA
ATATATCGCACCCGCATATATGCG
(iii)
TATATAGCGXXXGCGTATATA
ATATATCGCXXXCGCATATATGCG

Fig. 1C

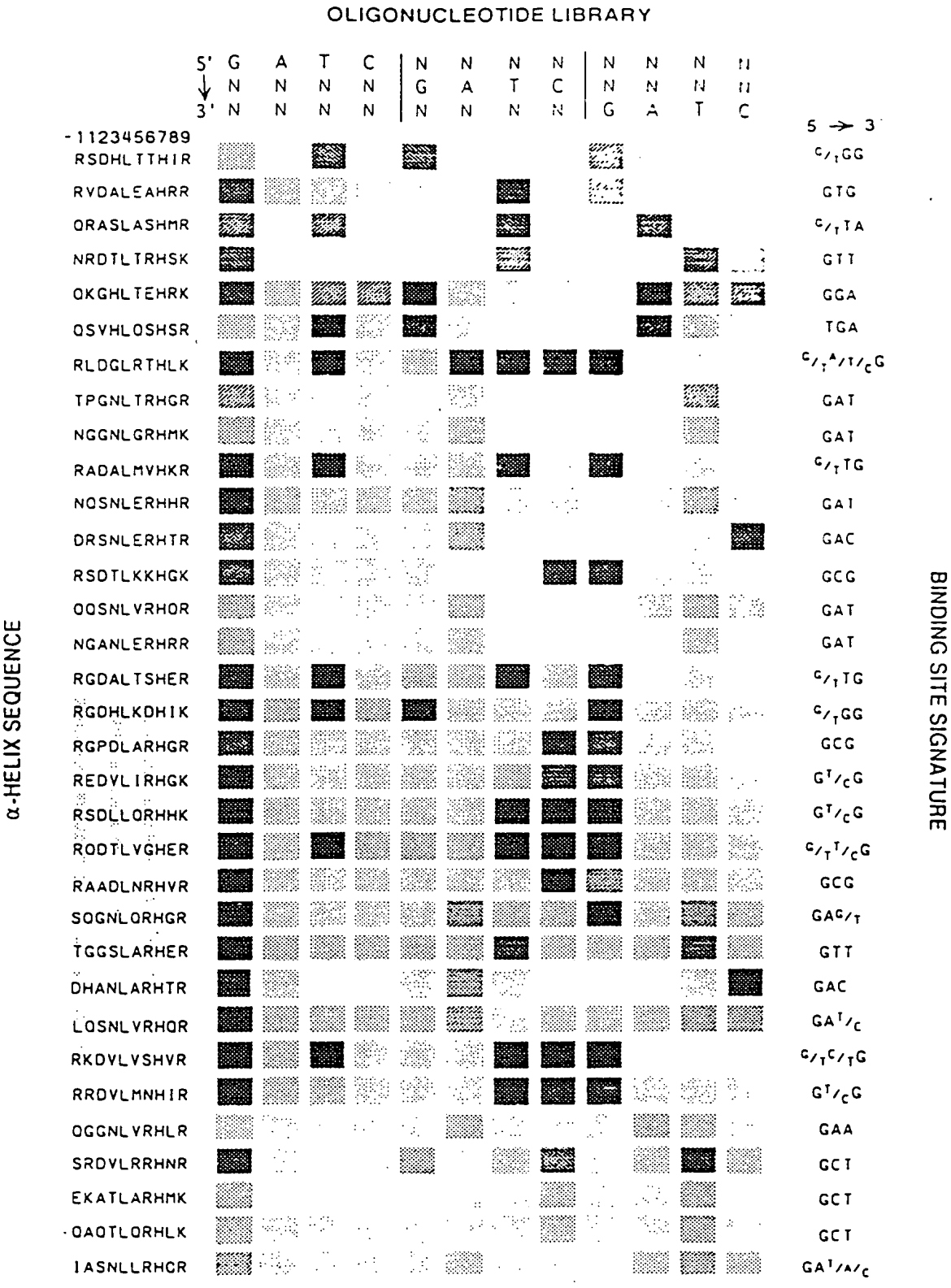

OLIGONUCLEOTIDE LIBRARY
5' G A T C | N N N N | N N N N
↓ N N N N | G A T C | N N N N
3' N N N N | N N N N | G A T C
-1123456789
5 → 3
RSDHLTTHIR
RVDALEAHRR
ORASLASHMR
NRDTLTRHSK
OKGHLTEHRK
OSVHLOSHSR
RLDGLRTHLK
TPGNLTRHGR
NGGNLGRHMK
RADALMVHKR
NOSNLERHHR
DRSNLERHTR
RSDTLKKHGK
OOSNLVRHOR
NGANLERHRR
RGDALTSHER
RGDHLKDHIK
RGPDLARHGR
REDVLIRHGK
RSDLLORHHK
RODTLVGHER
RAADLNRHVR
SOGNLORHGR
TGGSLARHER
DHANLARHTR
LOSNLVRHOR
RKDVLVSHVR
RRDVLMNHIR
OGGNLVRHLR
SRDVLRRHNR
EKATLARHMK
OAOTLORHLK
IASNLLRHGR
α-HELIX SEQUENCE
BINDING SITE SIGNATURE

FIG 2

FIG 3a

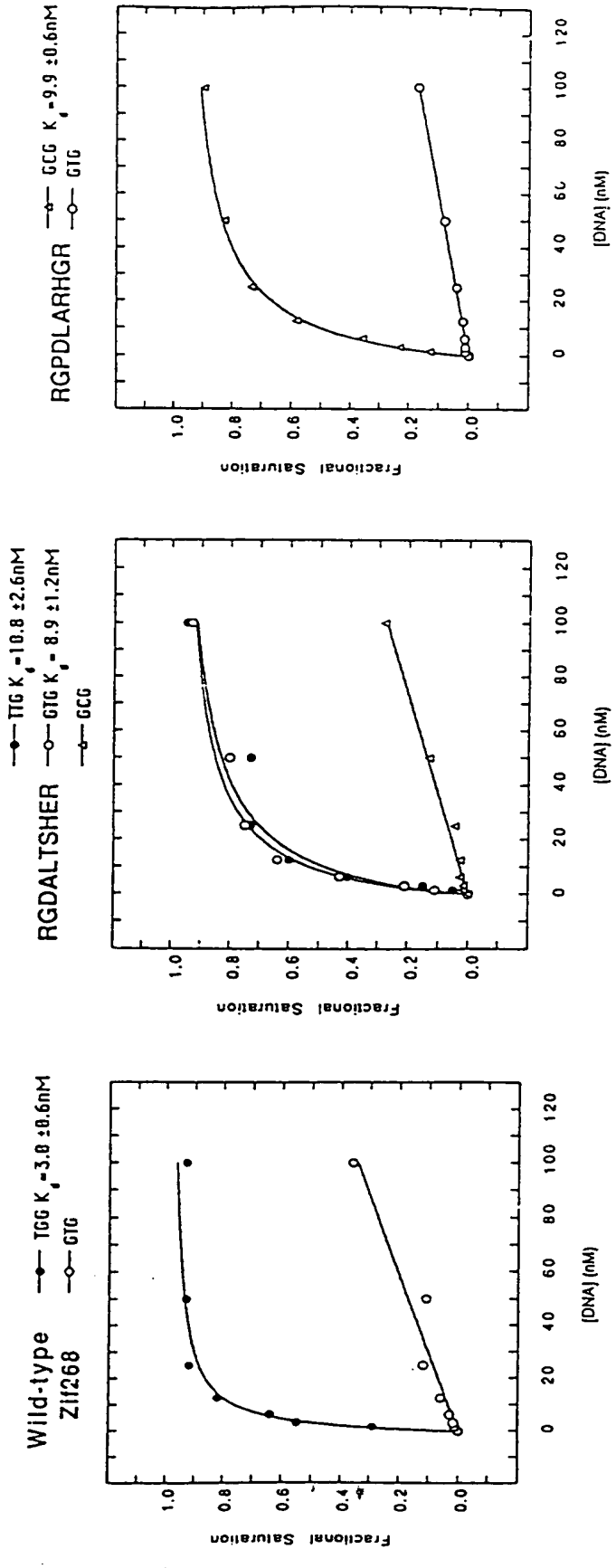
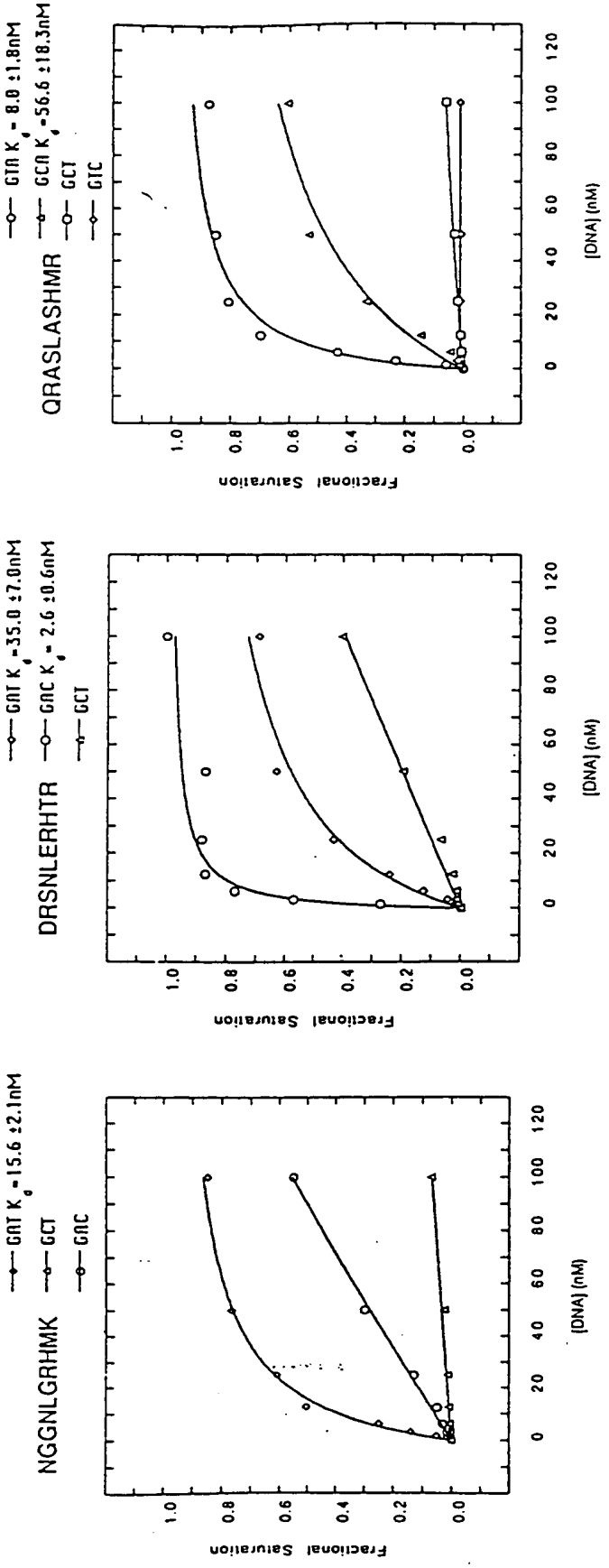

Wild-type
Zif268
TGG K_d =3.8 ±0.6nM
GTG
RGDALTSHER
TTG K_d =10.8 ±2.6nM
GTG K_d = 8.9 ±1.2nM
GCG
RGPDLARHGR
GCG K_d =9.9 ±0.6nM
GTG
Fractional Saturation
[DNA] (nM)
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120

FIG 3b


NGGNLGRHMK
GAT K_d=15.6 ±2.1nM
GCT
GAC
DRSNLERHTR
GAT K_d=35.0 ±7.0nM
GAC K_d= 2.6 ±0.6nM
GCT
QRASLASHMR
GTA K_d= 8.0 ±1.8nM
GCA K_d=56.6 ±18.3nM
GCT
GTC
Fractional Saturation
[DNA] (nM)
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120

1	*BCR - ABL*	TTC CAT GGA GAC GCA G AA GCC CTT CAG CGG CCA
2	*BCR*	TTC CAT GGA GAC GCA G gt gag ttc ctc acg cca
3	*ABL*	ccc ctt tct ctt cca g AA GCC CTT CAG CGG CCA

FIG 4

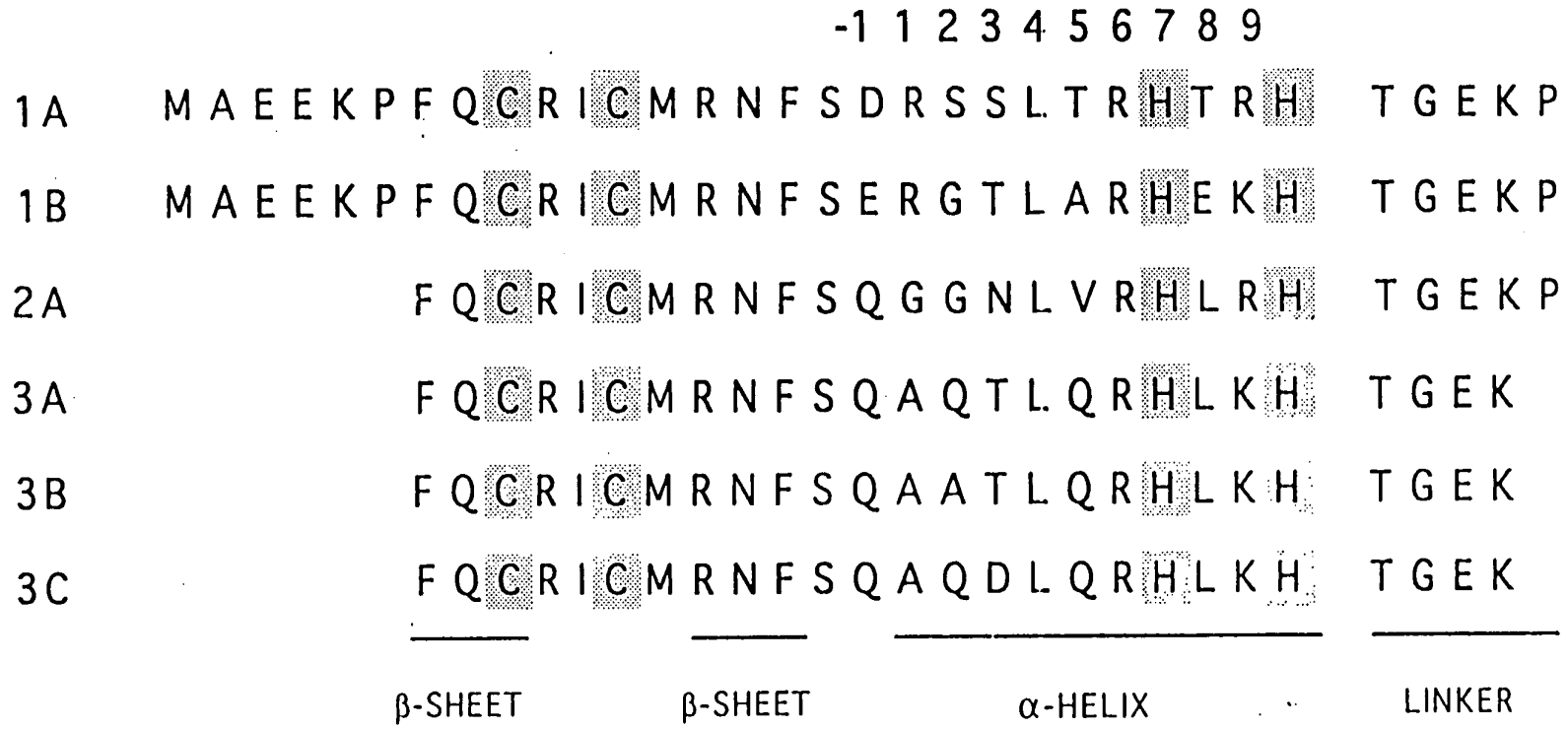
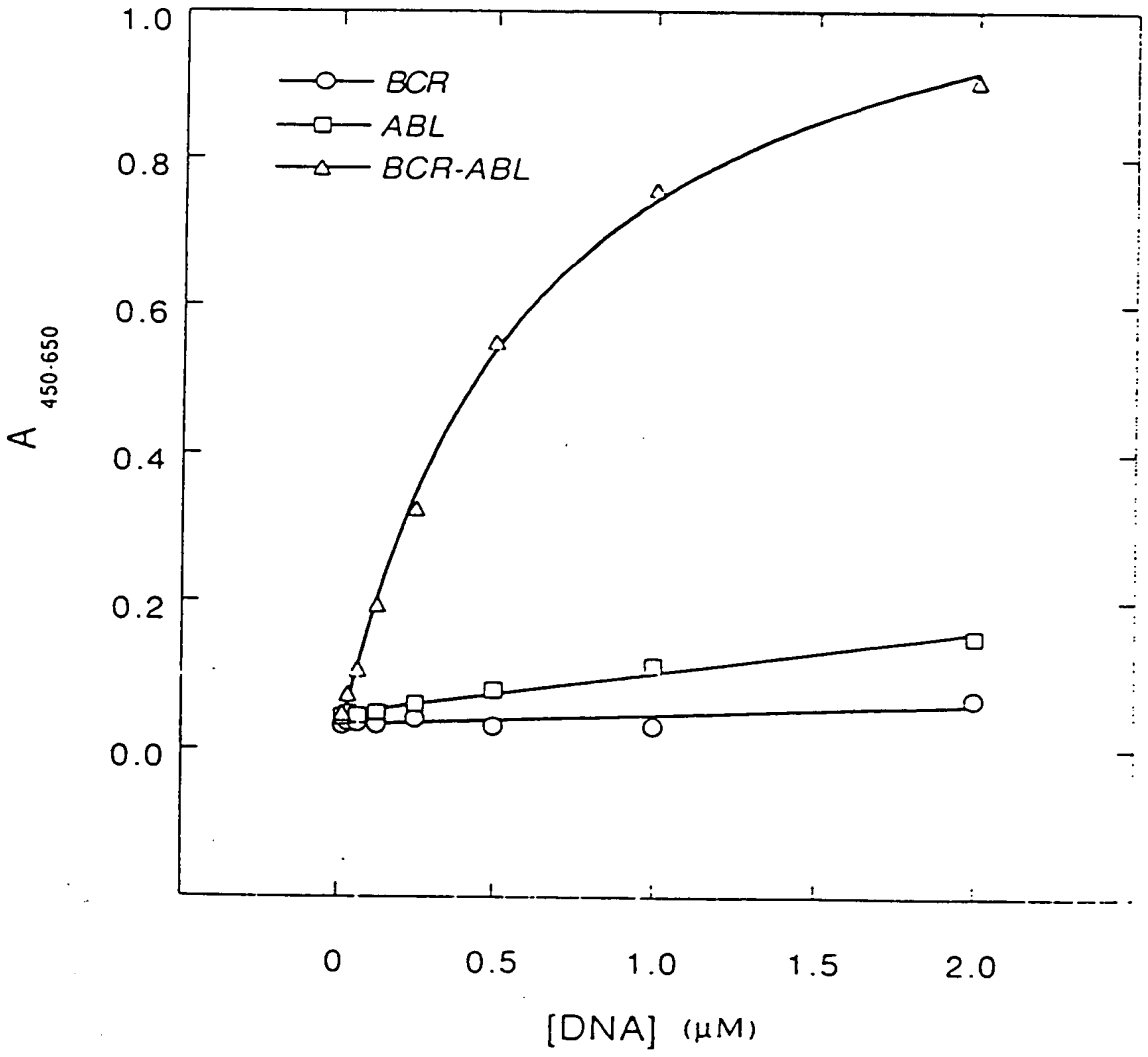
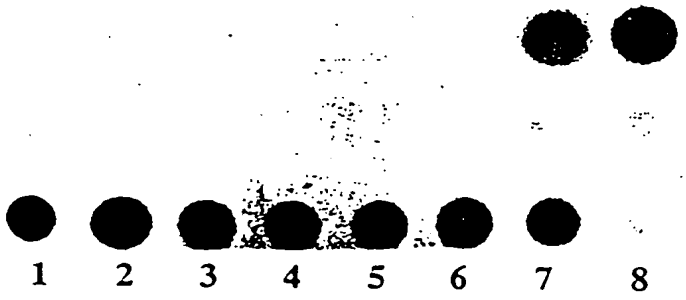

-1 1 2 3 4 5 6 7 8 9
1A M A E E K P F Q C R I C M R N F S D R S S L T R H T R H T G E K P
1B M A E E K P F Q C R I C M R N F S E R G T L A R H E K H T G E K P
2A F Q C R I C M R N F S Q G G N L V R H L R H T G E K P
3A F Q C R I C M R N F S Q A Q T L Q R H L K H T G E K
3B F Q C R I C M R N F S Q A A T L Q R H L K H T G E K
3C F Q C R I C M R N F S Q A Q D L Q R H L K H T G E K
β-SHEET
β-SHEET
α-HELIX
LINKER

FIG 5

FIG 6




1
2
3
4
5
6
7
8

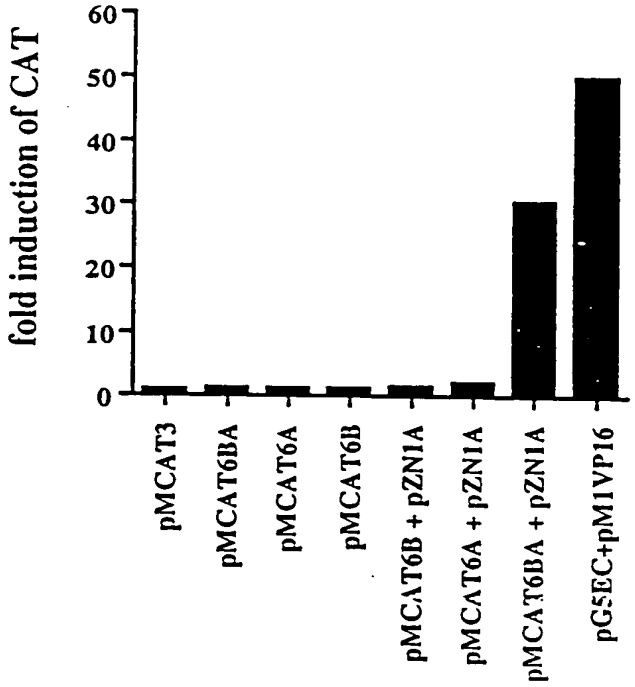
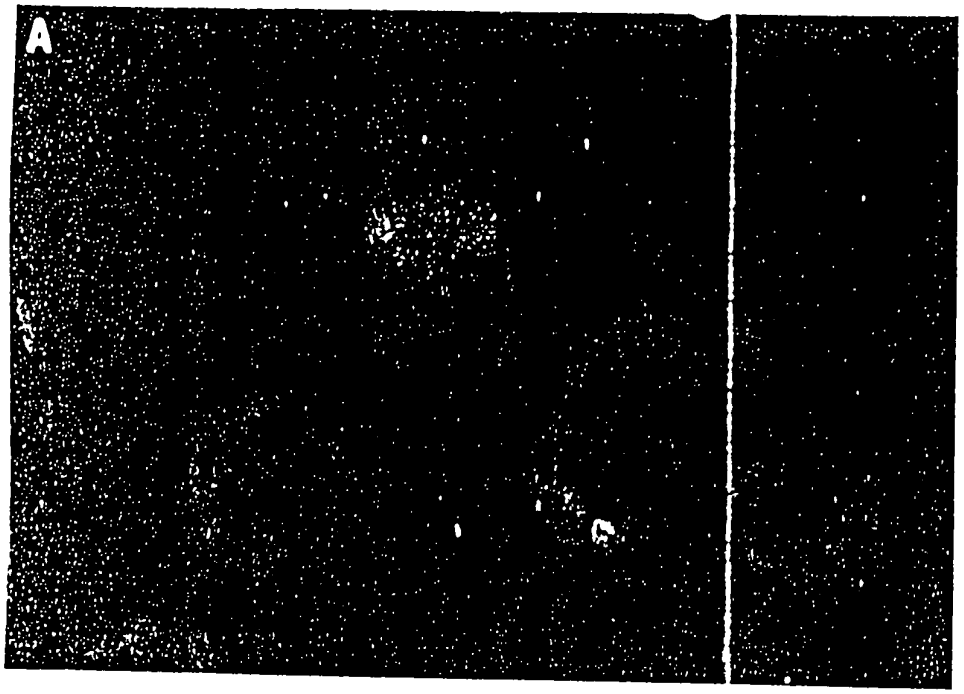
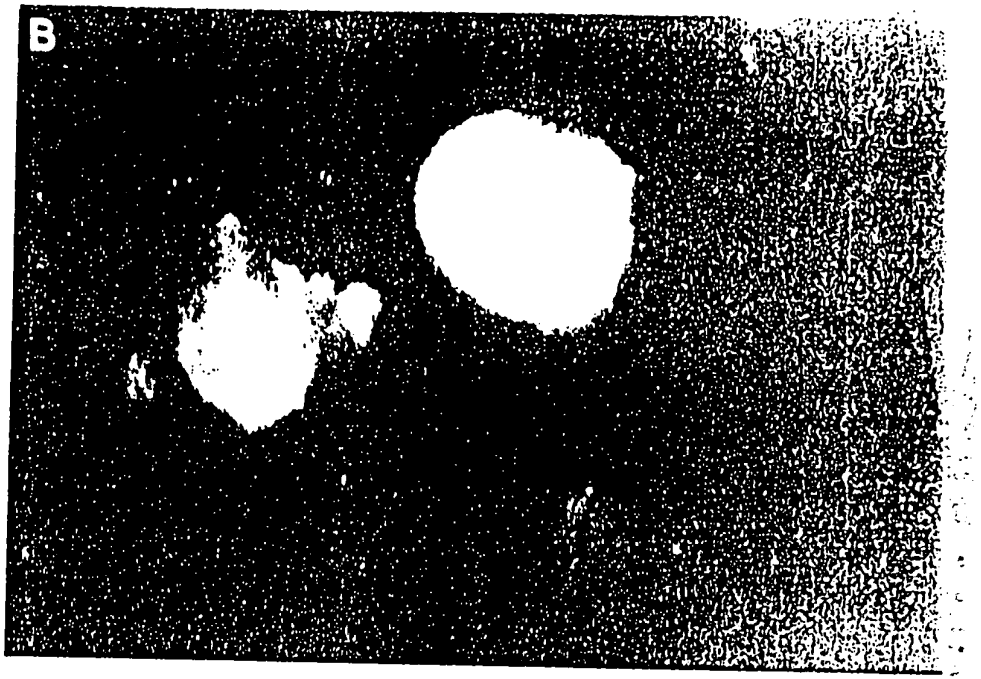
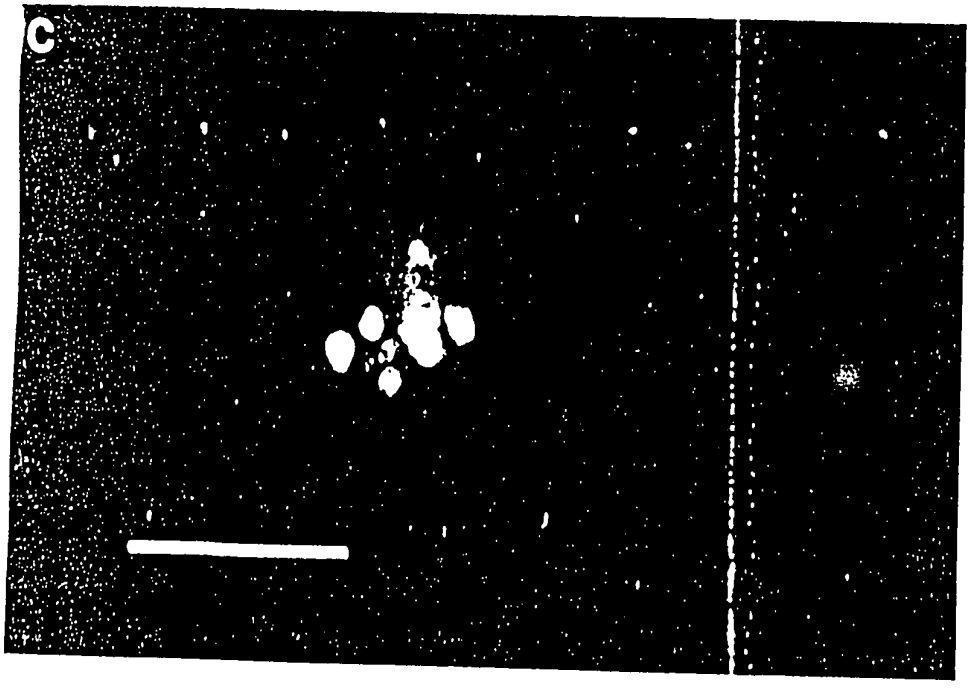

fold induction of CAT
60
50
40
30
20
10
0
pMCAT3
pMCAT6BA
pMCAT6A
pMCAT6B
pMCAT6B + pZN1A
pMCAT6A + pZN1A
pMCAT6BA + pZN1A
pG5EC+pM1VP16

FIG 7







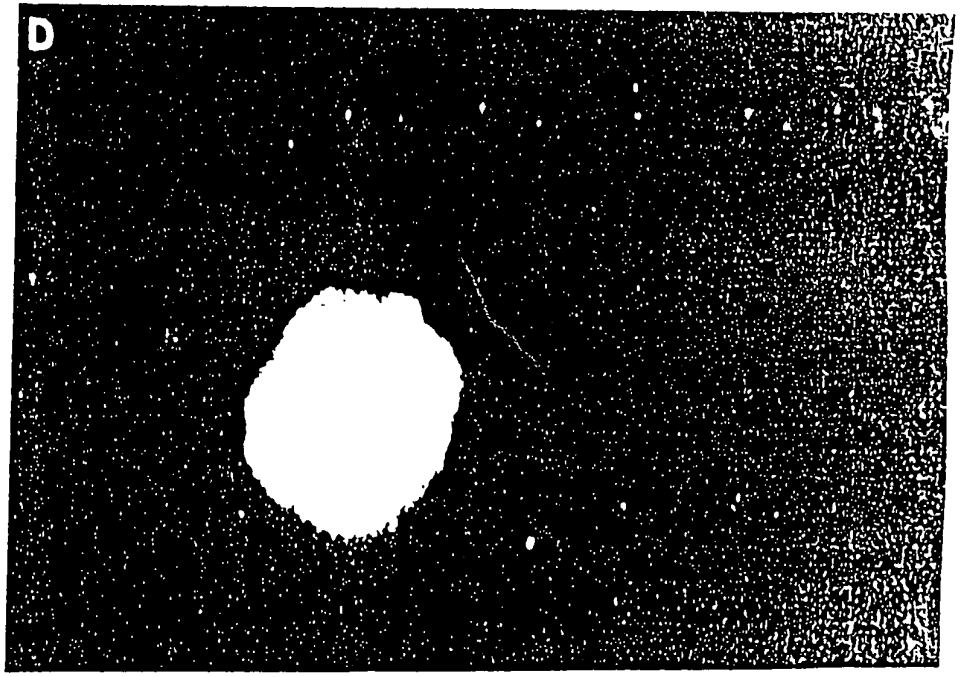


FIG 8

FIG 9A

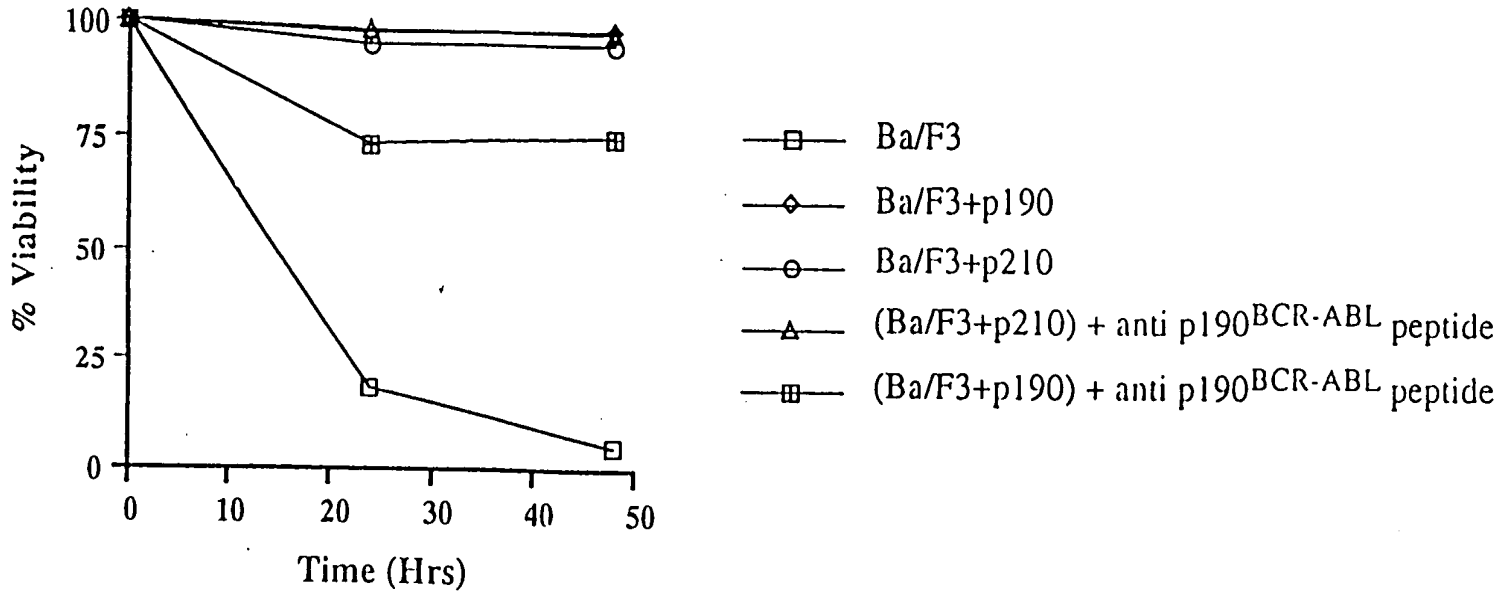

100
75
50
25
0
% Viability
0
10
20
30
40
50
Time (Hrs)
Ba/F3
Ba/F3+p190
Ba/F3+p210
(Ba/F3+p210) + anti p190BCR-ABL peptide
(Ba/F3+p190) + anti p190BCR-ABL peptide

Fig 9b